TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

PETTERI NYRHILÄ

IMPROVING MASTER DATA QUALITY IN DATA MIGRATION OF ERP IMPLEMENTATION PROJECT

Master of Science Thesis

Examiner: Professor Samuli Pekkola

Examiner and topic approved by the Faculty Council of the Faculty of Business and Built Environment on 3th December 2014

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Information and Knowledge Management

PETTERI NYRHILÄ: Improving master data quality in data migration of ERP implementation project
Master of Science Thesis, 103 pages, 5 Appendix pages
March 2015
Major: Business Information Management
Examiner: Professor Samuli Pekkola
Keywords: master data quality, data migration, SAP implementation

ERP systems constitute the information system backbone of most organizations across all industries. An ERP implementation is a huge commitment for organizations and too often they are failed or ran over schedule and budget. Poor data is claimed to be the number-one reason for the high failure rate of new computer system implementations. The problem is that the data which is migrated from a legacy system to a target system has poor quality.

The aim of this thesis was to study how the master data quality can be improved in the data migration process of SAP ERP implementation. The research was conducted at a general level and specific cases or organizations were not examined. The objective was to compile a list of methods how to improve the master data quality in the data migration process. To achieve the goal the main barriers for the master data quality in the data migration process were also recognized.

The research was conducted in two parts: theoretical and empirical. The theoretical section was based on scientific literature about the research title and formed the foundation for the empirical part. The empirical research was conducted as a case study where the qualitative data was collected by interviewing nine SAP consultants. Furthermore, an additional questionnaire was conducted in order to point out the most intrinsic results.

The results show that the data migration process includes several data quality barriers which need to be taken into consideration. The barriers were divided into three groups: data, people, and process related barriers. The methods to improve the master data quality in the data migration process were derived from the data quality barriers and they were also divided into the same three groups: data, people, and process relater methods.

According to the results the most intrinsic methods to improve the master data quality in the data migration of SAP implementation are taking care of the good engagement with the client, defining and communicating the data related roles and responsibilities unambiguously, analyzing the status of the data in the legacy systems at starting point, arranging and executing the data cleansing carefully already in the legacy system, creating unambiguous data collection templates and carrying out walkthrough for them, and determining SAP rules for data to correspond to the business rules.

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Tietojohtamisen diplomi-insinöörin tutkinto-ohjelma

PETTERI NYRHILÄ: Master datan laadun parantaminen toiminnanohjausjärjestelmän käyttöönoton datamigraatiossa
Diplomityö, 103 sivua, 5 liitesivua
Maaliskuu 2015
Pääaine: Tiedonhallinta
Tarkastaja: professori Samuli Pekkola
Avainsanat: master datan laatu, datamigraatio, SAP-käyttöönotto

Toiminnanohjausjärjestelmät muodostavat toimialasta riippumatta tietojärjestelmien perustan useimmissa organisaatioissa. Toiminnanohjausjärjestelmän käyttöönotto on organisaatioille iso sitoumus, ja liian usein ne epäonnistuvat tai menevät yli suunnitellun aikataulun ja budjetin. Huonon datan on sanottu olevan suurin syy käyttöönottoprojektien epäonnistumiselle. Ongelmana on, että data, joka siirretään vanhasta järjestelmästä uuteen järjestelmään, on huonolaatuista.

Tämän diplomityön tavoite oli tutkia, miten master datan laatua voidaan parantaa SAP-toiminnanohjausjärjestelmän käyttöönoton datamigraatiossa. Tutkimus suoritettiin yleisellä tasolla eikä työssä tutkittu tarkoin määrättyjä organisaatioita tai tapauksia. Tavoitteena oli muodostaa lista menetelmistä, joilla master datan laatua voidaan parantaa datamigraatioprosessissa. Tavoitteen saavuttamiseksi myös datamigraation oleellisimmat haasteet master datan laadulle tunnistettiin.

Tutkimus suoritettiin kahdessa osassa: teoriakatsauksena ja empiirisenä tutkimuksena. Teoria perustui aiheeseen liittyvään tieteelliseen kirjallisuuteen ja muodosti pohjan empiiriselle tutkimukselle. Empiirinen osuus suoritettiin tapaustutkimuksena, jossa laadullinen aineisto kerättiin haastattelemalla yhdeksää SAP konsulttia. Lisäksi kyselyä käytettiin täydentävänä menetelmänä osoittamaan olennaisimmat tulokset.

Tulokset osoittavat, että datamigraatiossa on useita erilaisia haasteita datan laadulle. Tässä tutkimuksessa ne jaettiin kolmeen ryhmään: dataan, ihmisiin, ja prosesseihin liittyviin haasteisiin. Menetelmät master datan laadun parantamiselle johdettiin tunnistetuista datan laadun haasteista, ja myös ne jaettiin kolmeen ryhmään: dataan, ihmisiin ja prosesseihin liittyviin menetelmiin.

Tulosten perusteella olennaisimmat menetelmät master datan laadun parantamiselle SAP-käyttöönoton datamigraatiossa ovat hyvän työsuhteen varmistaminen asiakkaan kanssa, dataan liittyvien roolien ja vastuiden määrittäminen ja kommunikoiminen yksiselitteisesti, datan tilan analysoiminen vanhassa järjestelmässä projektin alussa, datan putsauksen suorittaminen huolellisesti jo vanhassa järjestelmässä, yksiselitteisten datan keräykseen käytettävien pohjien luominen ja perehdyttäminen niiden käyttöön sekä määrittämällä säännöt datalle SAP:ssa siten, että ne vastaavat liiketoiminnan sääntöjä.

PREFACE

After months of work I can finally say, "It is done." The thesis project required several months, a number of weekends and long evenings spent planning and executing the research without mentioning the actual writing process.

In retrospect, the project itself was very interesting and I learned a lot about doing research projects in general. However, the most I have learned about the research topic itself, which also motivated me during the research to run with the project. I can honestly say that after the project I feel more confident about by competence regarding the research area than before.

I'd like to express my gratitude for all the people who helped me with the thesis project. Firstly, special thanks to the supervising Professor Samuli Pekkola for the professional advices and guidance during the research.

Also special thanks to Camilla and others who provided the opportunity for the thesis and supported me during the project. In addition, I'm grateful to all the interviewees for using their valuable time providing and sharing their experiences and ideas for the research.

Especially I'd like to thank Nanna who has been there for me and supported me during the whole project without mentioning the time before it. In addition, thank you Jenni for helping me to finalize the thesis. I also want to thank my whole family who has always supported me with everything.

Finally, my gratitude to all of my friends and especially to Roope and Samu for the support and the time we have spent either studying or having a leisure time apart from the studies.

Helsinki, 21th March 2015

Petteri Nyrhilä

TABLE OF CONTENTS

KEY TERMS AND ABBREVIATIONS

BAPI	Business application programming
Data migration	Process of selection, preparation, extraction, transformation and permanent movement of appropriate data from the legacy system to the target system.
ERP	Enterprise Resource Planning
ETL	Extract, transform, load
IDoc	Intermediate document
Legacy system	The original source system of the data
LSMW	Legacy system migration workbench
Master data	The core data and the main component of most information systems. It is created once, used many times and does not change frequently.
SAP	The target ERP system in this thesis
SD	Sales & distribution
Source system	Same as a legacy system, i.e. the original source system of the data
Target system	The final destination system of the data
XML	Extensible markup language

1. INTRODUCTION

Smart managers understand that the future of their organization depend on data and information. Data is crucial to everything they do from understanding customers to developing products and setting strategic directions. Savvy managers also sense that data and information are not just the lifeblood of the Information Age but it is a competitive advantage and the means by which they separate themselves from the competitors. Always when dealing with data and information the stakes are so high that they need to be treated as strategic assets. (Redman 2008, p. 1)

This thesis studied the data from ERP implementation project perspective. The main interest was to understand how the data quality can be improved in the data migration of an ERP system implementation. Firstly, this chapter provides a short background for the study. Secondly, the research problem, questions, objectives and focus are introduced. After that the research's methodological approach is explained. Lastly, the structure of the research is introduced.

1.1 Background

During the last couple of decades many firms' information technology strategy has followed the approach of implementing an organization wide enterprise resource planning (ERP) system as their most strategic computing platform (Hong & Kim 2002, p. 25). In a matter of fact ERP systems have been said to be one of most widespread IT solutions in the world (Motwani et al. 2005, p. 529-530). Even nowadays, an ERP system constitutes the information system backbone of most organizations across all industries (Kurbel 2013, p. 2).

ERP systems are cross industry systems supporting all major business processes within different kind of companies (Kurbel 2013, p. 2). ERP systems are designed to integrate and optimize various business processes such as order entry and production planning across the whole firm. An ERP system can integrate information used by accounting, manufacturing, distribution and human resources departments, etc. into a seamless computing system and when working successfully it can be the hearth of the organization's business intelligence and operative processes. (Motwani et al. 2005, p. 529-530.)

Nevertheless, history has taught us that too often information system implementation projects are doomed to failure (Hong & Kim 2002, p. 25; Kanaracus 2011; Kanaracus

2013). Poor data is claimed to be the number-one reason for the high failure rate of new computer systems (Redman 2008, p. 48). According to Kanaracus (2014) this is what happened for example when National Grid, a gas utility in New York, implemented a new SAP payroll system. Apart from the overly ambitious design and ineffective training, the reason why the project stumbled was poor data quality in the legacy systems. (Kanaracus 2014.) In addition, even if the poor data quality does not cause the failure of the implementation project it will have a major effect on the business as usual activities after the system is implemented.

Marsh (2005, p. 106) emphasizes that business data is the key asset that needs to be managed to ensure it is accurate, current and fit for purpose, i.e. high quality. Poor quality data has a direct impact on stocking levels, sales orders, customer perceptions, loyalty, profitability etc. It should be remembered that if the data is incorrect, irrelevant, incomplete, ambiguous or inconsistent it can distort the true picture and lead to poor and costly business decisions. (Marsh 2005, p. 106.) Thus, the consequences of the poor data quality can be huge. At the end of the day it can also lead to poor user experience and lower customer satisfaction, increased operating costs, inefficient decision-making process and lower performance. As the data is a critical input for almost all decisions it is impossible to make reliable and good decisions if the data quality is bad. (Haug & Arlbjørn 2011)

It has been stated that high data quality is important for business decision. However, managing data quality has been treated as a relatively low priority activity during past years (Marsh 2005, p.106). It has also been mentioned that even though ERP systems are widely used all over the world the implementation projects are too often failed. The challenge for these kinds of projects is often the data migration process (Marsh 2005, p. 107). The data migration means the process of selection, preparation, extraction, transformation and permanent movement of appropriate data from the legacy system to the target system (Morris 2012). Thus, the data migration plays obviously a critical role in the ERP implementation project from a data quality point of view.

As ERP systems have been the hot potato of the business world many years it is obvious that the title is widely studied (e.g. Monk & Wagner 2009; Phillips 2012; Wagner 2013). The same applies to master data management which aims to good governance and quality of the master data (e.g. Loshin 2009; Looi 2009; Roebuck 2012). However, fewer publications exist related to data migration. Even though the related titles are well-known, the title of this thesis in its respective context, data quality in data migration, is rather fresh approach to the data quality research. In addition, data is the foundation for operational, tactical and strategic decisions of the companies, i.e. the high-quality data is crucial to a company's success (Madnick et al. 2004). This fact makes the data quality an interesting title to study.

1.2 Research problem and research questions

The research problem was related to the phenomenon which is data migration as part of an ERP implementation project. This phenomenon was examined especially from the data quality point of view. The problem behind the study was that too often the data which is migrated from a legacy system to a target system has poor quality. Poor quality can mean for example incomplete data, invalid data or duplicate data records. In addition, usually the data models between the legacy system and the target system are inconsistent, which means that good quality data in the legacy system does not necessarily mean good quality data in the target system.

In any case poor data quality has a major effect on business due to the fact that it can cause unexpected and surprisingly high costs. For this reason it is essential that data quality is secured and preferably improved in the data migration process.

The primary research question is defined as:

- How to improve the master data quality in the data migration of SAP ERP system implementation?

The supporting secondary questions are:

- What is meant by good data quality?
- What is master data?
- What is data migration as a part of ERP implementation?
- What are the barriers for the data quality in the data migration process?

1.3 Research objectives and research focus

The main interest of this study was master data quality. The context where master data quality was studied is a data migration process as part of an ERP implementation. The objective of this thesis was to get recommendations how to improve master data quality in a data migration process. To achieve that goal main barriers for master data quality in a data migration process were also recognized.

The research was conducted at a general level and specific or unique cases or organizations were not examined. The purpose was not to create specific tools, models or frameworks but to compile a general list of methods how to improve master data quality in the studied phenomenon. The study was made thinking of the organizations who are about to renew their ERP system and are concerned about master data quality.

Data is very multi-valued term which is why the research objective needed to be more focused. Thus, the research focused only on master data and especially on sales and distribution (SD) related master data, such as customer data, material data and pricing related data. Other data types, such as transactional data and metadata, were excluded due to the fundamental differences between master data and other data categories and due to the fact that they may have different purpose, requirements and challenges, which directly affect the results how to improve data quality. However, the other data types were also shortly introduced in order to understand the difference between them. For the same reasons the other data objects, like vendor data or financial data, were excluded as well.

On the other hand, data migration is also a very wide title which is why this research focused only on the data migration of an ERP system implementation project and especially on SAP ERP system implementation. The research concentrated on the process where data is migrated from any legacy system to SAP system. Thus, SAP system can be described as a target system in this thesis.

On the contrary, what was not covered in this research is how the complete data migration process is exploited. To be precise this thesis concentrated on the data migration process only from data quality point of view and for example decommissioning the legacy data stores was not covered at a detailed level in this thesis. However, the complete process was described at general level in order to understand the overall picture. The purpose was not either to examine the master data management process and how data quality is managed at business as usual level. The studied phenomenon was clearly related to information system renewal process and especially ERP system implementation as described earlier.

One fundamental focus in this study was also the fact that the empirical part studied only the consultants' perspective. The customers' point of view was not studied at all due to the limitations of the tight schedule. However, the consultants' perspective provided skilled and in-depth insights into the ERP implementation project and how the data should be managed in it.

1.4 Research methodology

Before the actual research can be commenced, the researcher must consider questions which are philosophical by nature. This requirement does not apply only to theoretical or philosophical studies but also empirical studies. Even highly practical and real working life related studies are based on numerous hidden assumptions including e.g. human, world or information acquisition related presumptions. (Hirsjärvi et al. 2005, p. 129.) How the research questions are understood and how the research is designed are

clearly shaped by how the researcher views the world, his or her assumptions about human knowledge and about the nature of the realities encountered (Saunders & Tosey 2012, p. 58). Thus, it is considered highly important that before the actual research philosophical basis must be understood (Hirsjärvi et al. 2005, p. 129).

The figure 1.1. describes how the research philosophies, research approaches, research strategies and data collection and analysis methods are related to each other. It also illustrates which approaches and methods were used in this thesis.

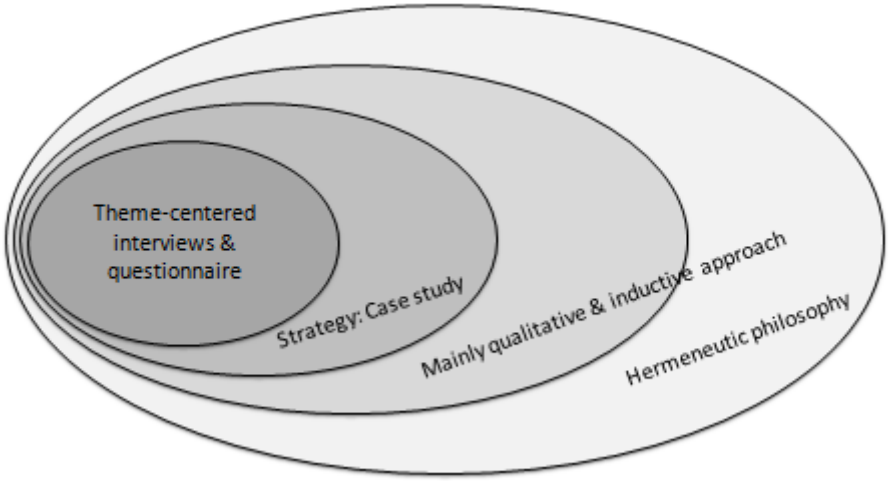

Figure 1.1. *The relation of research philosophies, approaches, strategies and data collection and analysis methods (adapted from Saunders et al. 2009).*

This chapter describes the three outermost sections of the figure 1.1. and justifies the chosen approaches. Firstly, the chosen research philosophy is explained. Secondly, the research approach is described and lastly the chosen research strategy is clarified. The data collection and analysis methods are explained at more detailed level in the chapter four.

1.4.1 Research philosophy

Saunders et al. (2009, p. 108) introduce four different research philosophies: positivism, realism, interpretivism and pragmatism. Typical for positivism is a researcher, e.g. a laboratory scientist, who is interested in observing and predicting results and concerned with law-like generalizations such as cause and effect relation. Realism emphasizes that what the researcher senses is the truth although senses are highly influenced by the researcher's world view and own experiences. On the contrary, the interpretive researcher realizes that what is being studied is a function of a specific set circumstances and individuals at a specific time. Finally, characteristic for pragmatism

is that the researcher considers that a single viewpoint cannot give the entire picture and that there may be several realities. (Saunders & Tosey 2012, p. 58).

Nevertheless, Olkkonen (1994, s. 26) states that the most considerable and widely used research philosophies especially for business economics related studies are positivism and hermeneutics. According to Heraclous (2004, p. 180) hermeneutics is considered as an interpretive approach. Butler (1998, p. 285) emphasizes that studies on information systems development have indicated that an interpretive approach is probably the most suitable for information system related phenomena. Butler (1998, p. 285) also states that among the several interpretive approaches hermeneutics has been described to be a valid interpretive approach for research on the phenomenon of the information systems development.

Olkkonen (1994, p. 35) describes that hermeneutic approach emphasizes understanding of the phenomenon and the research data is usually qualitative. Hermeneutic research philosophy also highlights subjective interpretations in the research. As Yin (2011, p. 310) states, hermeneutics is the aspect of a study that involves interpreting the event being studied to deepen the understanding of the political, historical, sociocultural and other contexts of the event. Consequently, it can be described as an opposite approach for those approaches which emphasize objectivity.

It can be concluded that research philosophies can be divided in many different groups and there are fundamental differences between different philosophies and how they are understood in different schools. As this research was related to information system development the most appropriate approaches for the research philosophy were interpretive approaches. Thus, this thesis followed mainly hermeneutic philosophy as the target of this research was to understand the context of the studied phenomena. The results of the research were also more or less subjective considering the fact that the volume of the empirical data was limited. In addition, the main research data was also qualitative in nature which suits hermeneutics. All of these facts support the conclusion that hermeneutics was an appropriate philosophy for this research.

1.4.2 Research approach

Research approaches can be often divided to different counterparts. One widely used style to describe research approaches is to divide those to qualitative and quantitative approaches (Hirsjärvi et al. 2005, p. 136). According to Saunders et al. (2009, p. 151) the terms quantitative and qualitative are used widely in business and management research to differentiate both data collection techniques and data analysis procedures.

Quantitative is mainly used as a synonym for any data collection technique, such as questionnaire, or data analysis procedure that generates or uses numerical data

(Saunders et al. 2009, p. 151). It is also based on exact and thorough planning when it comes to choosing test subjects and sample sizes. The studied phenomenon has to be set to a specific theory and hypotheses are often presented. The conclusions in a quantitative research are also based on statistical analysis. (Hirsjärvi et al. 2005, p. 140.)

Qualitative is used basically as a synonym for any data collection technique, such as interviews or observation, or data analysis procedure that generates or use non-numerical data (Saunders et al. 2009, p. 151). In the qualitative research humans are favored as instruments of collecting the data. This means that the researcher trusts more on his/her own observations and discussions with the people than questionnaire based data collection methods. Usually the research is also based on the fact that the cases are unique and the research material is interpreted keeping that fact in mind. Often inductive analysis is used in a qualitative research. (Hirsjärvi et al. 2005, p. 164.)

However, Alasuutari (2001, p. 32) emphasizes that qualitative and quantitative research approaches are not the opposites or do not rule out one another. Both qualitative and quantitative approach can be utilized even within the same research (Alasuutari 2001, p. 32). Also Hirsjärvi et al. (2005, p. 136) state that quantitative and qualitative researches can be used as complementary approaches for each other. In this thesis the primary research data was qualitative in nature and the research was based on the fact the cases are unique. Thus, the main data collection technique was qualitative (as it will be described in the chapter four). However, based on the qualitative data collection results also a quantitative data collection technique was used as an additional technique to bring extra insights into the research data. Therefore, it can be stated that the main approach used in this thesis was qualitative but there were also quantitative features.

Yin (2011, p. 93) describes another way to divide research approaches to inductive and deductive approach, as these reflect different ways of shifting between data and concepts. Yin (2011, p. 94) states that inductive approaches let the data lead to the emergence of concepts whereas deductive approaches let the concepts lead to the definition of the relevant data.

An inductive approach is often used in a qualitative research (Yin 2011, p. 94). Hirsjärvi et al. (2005, p. 164) also defines that the purpose of an inductive research is not to test theories or hypotheses and the researcher does not choose what is important. On the contrary, in a deductive approach the analysis of the research data is based on existing theories or models (Tuomi & Sarajärvi 2002, p. 95). However, a deductive approach can save the researcher from suffering a lot of uncertainty when doing the initial field work on data collection as it helps to identify the relevant concepts beforehand. On the contrary, a major risk with this approach can be losing any fresh insights. (Yin 2011. p. 95.)

In this thesis the main purpose was not to test theories or hypotheses. This means that the analysis was not based on existing theories. From this viewpoint this thesis followed an inductive approach. However, a pure inductive approach is very hard to execute because the research findings are not necessarily totally independent from the researcher as for example the research methods might influence on the results and the methods are chosen by the researcher (Tuomi & Sarajärvi 2002, p. 98). Also in a successful inductive approach there is no theory or preconceptions which could affect the field work at the beginning point of the research (Yin 2011, p. 124).

Even though this research mainly followed an inductive approach the theory had its purpose as it defined the preliminary knowledge for the research. As the related titles, e.g. ERP implementation, data migration and master data management, are widely studied the initial theory for the research problem could be easily found and utilized. However, the purpose was not to limit too much the empirical part according to the theory but just to help the data collection by recognizing the high-level concepts in advance. As Saunders et al. (2009, p. 490) remind, while one may commence with either an inductive or a deductive approach in practice the research is likely to combine elements of both. This applied also in this study.

In other words, before the data collection the initial theory was recognized and studied. According to the theory the primary concepts were picked out for the empirical part of the research. From this viewpoint there are also some deductive features in this research. Nevertheless, the main stress in the research was in the empirical research data and the preconceptions during the fieldwork were outlined to minimum. Thus, during the empirical research it was not the intention to influence too much on the research data but to let the field speak for itself. Hence, the possible preconceptions raised from the initial theory could be overruled during the empirical research.

1.4.3 Research strategy

According to Hirsjärvi et al. (2005, p. 134) there are three traditionally used research strategies: experimental research, survey research and case study. An experimental research strategy usually measures how some variable affects some other variable. Typical for this is that a chosen sample is studied with different experiments when the research circumstances are changed systematically. An experimental research involves usually testing hypotheses. (Hirsjärvi et al. 2005, p. 134.)

Characteristic for a survey research is that the research data is collected from a group of people in a standardized format. The sample group is carefully planned from a certain group of people. Data is usually collected using a structured interview or a

questionnaire. (Hirsjärvi et al. 2005, p. 134.) However, Yin (2009, p. 34) states that surveys' ability to investigate the context is extremely limited.

On the contrary, typical for a case study is that it usually concentrates on processes and the cases are studied in touch with their environment and context (Hirsjärvi et al. 2005, p. 135). Yin (2009, p 33) also confirms that a case study investigates a contemporary phenomenon within its real-life context in depth. In addition, Hancock and Algozzine (2006, p. 16) define that the phenomenon being researched is studied in its natural context, bounded by space and time.

The peculiar need for case studies arises out of the urge to study and understand complex social phenomena. The case study allows researchers to retain the holistic and meaningful characteristics of real-life events, like organizational and managerial processes. (Yin 2009, p. 22.) The case study can concentrate on one unique case or cover multiple cases and then have a set of conclusions drawn from those (Yin 2009, p. 35). Hancock and Algozzine (2006, p. 15) also state that although case study research sometimes focuses on an individual representative of a group more often it addresses a phenomenon, e.g. a particular event, situation, program, or activity.

The studied case or phenomenon in a case study can be for example individual, family, community, organization, event or process (Aaltola & Valli 2007, p. 188). Hancock and Algozzine (2006, p. 15) also support that the topics of case study research vary widely. For example, case studies of programs, events, persons, processes, institutions, social groups, and other contemporary phenomena have been completed. (Hancock & Algozzine 2006, p. 15).

In a case study research the research material is often collected using variety of methods, like observing, interviewing or studying related documents (Hirsjärvi et al. 2005, p. 135). Saunders et al. (2009, p. 360) also emphasize that a case study research strategy can also make use of questionnaires. It is stated that a case study is relevant choice for getting answers to "what" and "how" questions. (Saunders et al. 2009, p. 146).

This thesis is answering the question "how". The main data collection method was interview, as it will be described in the chapter four, but also survey based questionnaire was used in addition to give extra insights. The purpose was to study a phenomenon which is data migration process where the main interest is master data quality. The phenomenon was studied in its natural context. The approach was to study the phenomena at a general level rather than in a specific circumstance. This means that the research data was not collected from a particular case but from the experiences of similar cases in general. The key factor was to understand the context where the data

migration phenomenon happens. Thus, it can be concluded that the used research strategy in this thesis was a case study.

1.5 Structure of the research

The chapter one forms the introduction of the thesis. Apart from the introduction this thesis includes four main parts: theory, empirical study, results, and discussion. In addition, the final part is summary.

The chapters two and three form the theory of the thesis. The chapter two discusses about the data quality and master data quality. In addition, it forms the basic theory for the research. The chapter three discusses about ERP implementations and especially data migration processes within the implementation projects. The theory ends to the discussion about barriers in the data migration process.

After the theory the chapter four introduces the empirical study. It describes the data collection and analysis methods and explains how the empirical part of the research was conducted.

The chapter five describes the results of the empirical research. The results are discussed in four subchapters so that the first subchapter discussed about the data quality and perfect data quality, the second about the data quality barriers in the data migration, the third about improving the master data quality in the data migration and the fourth about the prospects.

The chapter six forms the discussion about the results. The results are observed from different perspectives and reflected in contrast to the theory. The structure of this chapter follows the structure of the chapter five.

Finally, the chapter seven forms the conclusion of the research. Firstly, the results of the research are summarized in relation to the research questions. Secondly, the conducted study is assessed and the potential further research is discussed.

2. DATA QUALITY AND MASTER DATA

In order to discuss the concept master data quality it is essential to understand the concept of data and what is meant by data quality. Therefore, this chapter provides firstly the definitions for data and data quality. Secondly, in the second subchapter the concept of master data is discussed at more detailed level.

2.1 Data and data quality

As a basis for the discussion of data quality it is useful to start with the definition for data. Therefore, this chapter starts with the discussion about data and then moves forward to data quality and data quality problems.

2.1.1 Data, information and knowledge

At first, it may be useful to distinguish data from information and knowledge. It is said that knowledge is derived from information which is in turn derived from data. It is also claimed that wisdom can be derived from knowledge. (Redman 2008, p. 28.) Thus, it can be concluded that data has an important role when creating information and knowledge. The figure 2.1. illustrates the relation of data, information, knowledge and wisdom.

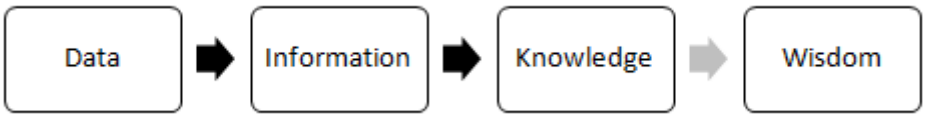

Figure 2.1. *The relation of data, information, knowledge and wisdom.*

Tayi and Ballou (1998, p. 56) claim that in a very real sense data constitutes the raw material for the Information Age. However, unlike physical raw material data is not consumed and in fact can be reused repeatedly for various purposes (Tayi & Ballou 1998, p. 56). In a theoretical sense Davenport and Prusak (1998, p. 2) define data as a set of discrete, objective facts about events. Again, in an organizational context data can be described as structured records of transactions (Davenport & Prusak 1998, p. 2). According to Haug et al. (2011, p. 171) Newell et al. (2002) define data as providing a record of signs and observations collected from various sources. Zack (1999, p. 46) on

the other hand provides the definition that data represents observations or facts out of context and therefore is not directly meaningful.

Redman (2008, p. 14) describes that in a business context the concept of data consist of two components: a data model and data values. Haug et al. (2009, p. 1055) and Redman (2008, p. 14) sum up that a data model is a definition of entities (specifications of the things of interest), their attributes (important properties of those things), and relationships between them. Data values complete the picture as they are assigned to attributes for specific entities (Redman 2008, p. 14). An example of an entity is "employee", which could have the attributes "ID", "name", "date of birth", "address", etc. The employee could have a relationship to the entity "ERP project", implying that the employee is involved in this project. Again the data value for the employee attribute "ID" could be "021546". When such data are stored they become data records. (Haug et al. 2009, p. 1055.)

Haug et al. (2009, p. 1055) claim that unlike other typical resources data can be copied, shared and combined in many ways. Thus, compared with other resources data has some special challenges to be aware of, e.g. to ensure that data is of adequate quality (Haug et al. 2009, p. 1055). Redman (2008, p. 12-29) concludes the wonderful and hazardous properties of organizational data (and information) as follows:

- Data multiplies
- Data is more complex than it appears
- Data is subtle and nuanced and has become the organization's lingua franca
- Data creates value when it is on the move
- Data is organic
- Data can be shared, lost, or stolen
- Data is not consumed with use
- Data is the mean by which organizations encode knowledge
- Data is intangible
- Each organization's data is uniquely its own

Davenport and Prusak (1998, p. 4) say that information can be thought as data that makes a difference. They define information as data transformed by the value-adding processes of contextualization, categorization, calculation, correction and condensation (Davenport & Prusak 1998, p. 4). On the other hand, Haug et al. (2011, p. 171) concludes that Newell et al. (2002) define information as data presented in a particular way in relation to a particular context of action. Again Zack (1999, p. 46) states that information results from placing data within some meaningful context and often in the form of a message.

Davenport and Prusak (1998, p. 5) state clearly that knowledge is not neat or simple concept to define. Thus, they describe that knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. They say that knowledge derives from minds at work and in organizations it often becomes embedded in documents, repositories, organizational routines, processes, practices, and norms. (Davenport & Prusak 1998, p. 5.) Zack (1999, p. 46) defines that knowledge is that which is believed and valued on the basis of the meaningful accumulation of information through experience, communication, or interference.

2.1.2 Data quality in general

Data quality is not a straightforward term to define. To design for better quality, it is necessary first to understand what quality means and how it can be measured (Wand & Wang 1996, p. 87). Therefore, it is essential to discuss the concept of data quality in more detailed level.

A classic definition of data quality is "fitness for use", i.e. the data quality is good when it successfully serves its purposes (Tayi & Ballou 1998, p. 54; Haug & Arlbjørn 2011, p. 292). In addition, what we call data may well be meaningless unless placed in some context (Tayi & Ballou 1998, p. 56). Thus, data quality can be seen as a contextual or relative concept (Haug & Arlbjørn 2011, p. 292). Tayi and Ballou (1998, p. 54) also emphasize that data with quality considered appropriate for one use may not possess sufficient quality for another use. On the other hand, Orr (1998, p. 68) defines data quality as the measure of agreement between the data views presented by an information system and that same in real world.

Pipino et al. (2002, p. 211) emphasize that studies have confirmed that data quality is a multidimensional concept (e.g. Ballou & Pazer 1985; Wand & Wang 1996; Wang & Strong 1996). Thus, another way to understand the concept of data quality is by dividing it into dimensions. McGilvray (2008, p. 7) claims that data quality dimensions are facets of data quality which can be used to measure and manage data and information quality.

In literature data quality has been divided into dimensions by various ways. Referring to the literature review by Wang et al. (1995) Wand and Wang (1996, p. 92) mention several notable data quality dimensions which are represented in the figure 2.2.

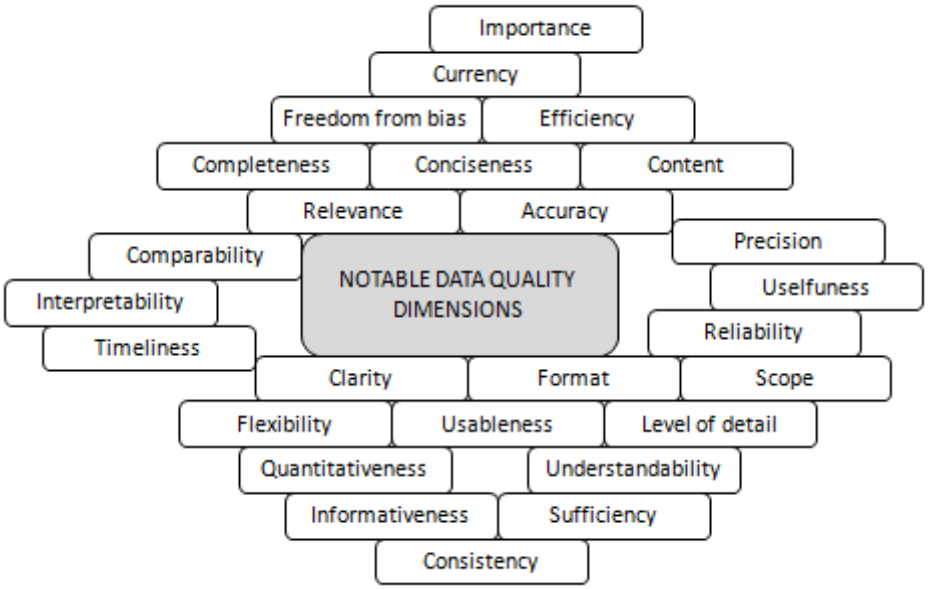

Figure 2.2. *Notable data quality dimensions (adapted from Wand and Wang 1996).*

Different data quality dimensions were recognized already in the 80's when Ballou and Pazer (1985, p. 153) introduced four dimensions for data quality: accuracy, timeliness, completeness and consistency. Wand and Wang (1996, p. 87) refer to these as frequently mentioned dimensions. Ballou and Pazer (1985, p. 153) emphasize that the most straightforward dimension is the accuracy and it basically means the difference between the correct value and that actually used. The timeliness is also a rather simple dimension as it states that if the stored value is outdated it is in error. The completeness means that all values should be recorded and no data is missing. The consistency dimension requires that the representation scheme should be standard for all values. (Ballou & Pazer 1985, p. 153.)

Based on their research Wand and Wang (1996, p. 92) define a model of four intrinsic data quality dimensions which are completeness, unambiguousness, meaningfulness, and correctness. These intrinsic dimensions differ from the earlier illustrated notable data quality dimensions. However, according to Wand and Wang (1996, p. 93) these intrinsic dimensions are related to many of the data deficiencies in the group of notable data quality dimensions. For example, inaccuracy and lack of precision are related to unambiguousness. Furthermore, reliability can indicate whether the data can be counted on to convey the right information, which is related to correctness. (Wand & Wang 1996, p. 93.)

Also Pipino et al. (2002) introduce a list of different data quality dimensions and their definitions. Many of these are also included on the group of notable data quality dimension illustrated earlier but not all. The table 2.1. combines the data quality

dimensions and their definitions introduced by Ballou and Pazer (1985), Wand and Wang (1996), and Pipino et al (2002).

Table 2.1. *Data quality dimensions and definitions (combined from Ballou & Pazer 1985; Wand & Wang 1996; Pipino et al. 2002).*

Data quality dimension	**Definition**
Accuracy	The recorded value is in conformity with the actual value
Unambiguousness	Insufficient information: the data can be interpreted in more than one way
Meaningfulness	It is not possible to interpret the data in a meaningful way
Accessibility	The extent to which data is available, or easily and quickly retrievable
Appropriate amount of data	The extent to which the volume of data is appropriate for the task at hand
Believability	The extent to which data is regarded as true and credible
Completeness	The extent to which data is not missing and is sufficient breadth and depth for the task at hand
Concise representation	The extent to which data is compactly represented
Consistent representation	The extent to which data is presented in the same format
Ease of manipulation	The extent to which data is easy to manipulate and apply to different tasks
Correctness	The extent to which data is correct and reliable
Interpretability	The extent to which data is in appropriate languages, symbols, and units, and the definitions are clear
Objectivity	The extent to which data is unbiased, unprejudiced, and impartial
Relevancy	The extent to which data is applicable and helpful for the task at hand
Reputation	the extent to which data is highly regarded in terms of its source and content
Security	The extent to which access to data is restricted appropriately to maintain its security
Timeliness	The extent to which data is sufficiently up-to-date for the task at hand

Understandability	The extent to which data is easily comprehended
Value-added	The extent to which data is beneficial and provides advantages from its use

Also many other authors have defined different data quality dimensions (e.g. Olson 2003; McGilvray 2008). The data quality dimensions help to understand what requirements the data needs to fulfill in order to be good quality. However, Eckerson (2002, p. 11) emphasizes that defect-free data is not required. It is nearly impossible to ensure that all data meet the criteria of perfect data quality. It simply needs to meet the requirements of the people or applications that use it. And different people, applications, and purposes of the data require different quality levels of the different data quality dimensions. (Eckerson 2002, p. 11.)

It can be concluded that data quality is highly multidimensional concept hence the data quality can be difficult to understand in its whole aspect. Nevertheless, as it was already mentioned it is highly useful to recognize the various data quality dimensions as they can be used to measure and manage data quality (McGilvray 2008, p. 7). Like Bobrowski et al (1999 p. 4) emphasize the well-known phrase: "if you can't measure it, you can't manage it". Thus, data quality can only be improved if it can be measured (McGilvray 2008, p. 7), which requires understanding the multidimensional concept of data quality.

2.1.3 Data quality problems

As it was proven data quality is a multidimensional concept. Thus, there are many ways for data to be wrong which makes ensuring data quality challenging task to do. To be realistic Tayi and Ballou (1998, p.56) state that no one can anticipate all the circumstances that could compromise the integrity of an organization's data. However, it is important to understand what kind of issues data can have to understand better where they are derived from.

According to Redman (2008, p. 41) the seven common data quality issues are as follows:

- People cannot find the data they need
- Incorrect data
- Too much data
- Data inconsistency across sources
- Poor data definition
- Data privacy and data security

- Organizational confusion

Redman (2008, p. 42) suggests that according to studies (e.g. Feldman & Sherman 2001; WorldOne Research 2008) knowledge workers spend up to 30% of their time searching for the data they need. This describes how hard it is to find the right data when it is necessary. As Feldman & Sherman (2001, p. 1) concludes, timely access to critical information separates the winners from the losers in today's information economy. In the fast-paced business environment one has to be able to access quickly on the right data and to be able to make fast decisions on the strength of it. If the data cannot be found it can result poor decisions, duplicated efforts, lost sales and lost productivity for the enterprise (Feldman & Sherman 2001, p. 5-6).

Incorrect data is a clear issue to understand as it means the data where the values disagree with their real-world counterparts. Redman (2008, p. 42) estimates that 10 to 25 percent of the data records contain errors including also missing data but it needs to be noted that there is a variation between different databases and organizations. Making decisions on the strength of the incorrect data can clearly lead to poor decisions. In addition, from another perspective the data can be valid even though it is incorrect, e.g. the social security number of a person can be valid but it is just not the right person's number (Eckerson 2002, p. 12).

Related to the correctness of the data also Tayi and Ballou (1998, p. 56) give an example concerning foreign exchange rates in today's newspaper which in a sense are timely and yet in actuality the values were out of date before the newspaper was printed. This affects the timeliness of the data. As another example by Tayi and Ballou (1998, p. 56) in a particular file every data value could be both accurate and timely but certain records may be missing entirely. Also the sales data may be accurate and timely and yet inconsistent and hence of little use if an inappropriate reporting period is used. (Tayi & Ballou 1998, p. 56.)

The amount of data is vastly increasing. The reason for this is quite simple. All activities that use data create more in the process. Every operation, every decision and every strategic action create more data, e.g. even taking a customer order creates new data. (Redman 2008, p. 12.) No wonder that organizations have too much data on their hands. Redman (2008, p. 41) also claims that half of all data is never used for anything. This causes time and money issues and keeps people from focusing on more important data.

Too much data leads also to data redundancy which can contribute to inconsistency. Even two copies of a data record, once identical, used by different departments can after a while be different, which confuses knowledge workers and complicates the collaboration between the departments. (Redman 2008, p. 43.) The root cause is the

fragmentation of organization into a multiple departments, divisions, and operating groups (Eckerson 2002, p. 13).

The lack of clear and commonly understood data definitions is also a critical issue for data quality (Redman 2008, p. 43). It can lead to misunderstandings and critical problems if the data definitions are not commonly shared and understood across the whole organization. In a fragmented organization slowly each group begins to use slightly different definitions for common data entities, such as customer data, and apply different rules for calculating values, such as net sales and gross profits (Eckerson 2002, p. 13).

Hackers have been after data for years (Redman 2008, p.44). Hence it is essential to secure the business critical data. However, when technology and data security is evolving also the hackers are becoming cleverer. Data privacy violations can also come in many forms and organizations need to be careful how they handle and utilize their own data in order not to cross any data privacy regulations.

The last but not the least point in Redman's list is organizational confusion. Even though most companies consider data as a business critical asset, they don't know how to answer the basic questions concerning their data. Not having answer to the questions, like what data do you have, where are they, which are the most important, how do you use them, where do they come from and what are they worth, exacerbates the other issues. For example, it is a waste of effort to improve the accuracy of data no one ever uses. (Redman 2008, p. 44-45.) It is a fact that the organizations do not necessarily know their data and especially the status of the data quality. In addition, some managers refuse to admit that their data are not good enough (Redman 2013, p. 6).

To conclude the data issues Redman (2008, p. 41) listed, too much data is just plain wrong, too hard to find, poorly defined, inconsistent with other data, and at risk of being lost or stolen. Organizations do not know what they have, redundancy is out of control, and too much data are never used for anything (Redman 2008, p. 45).

2.2 Master data quality

In order to understand the concept of master data it is useful to introduce also the other data categories. This is done in the first subchapter. Secondly, the sales and distribution related master data is introduced to understand the main focus area of the research. Lastly, the effects of poor master data quality are discussed.

2.2.1 Master data definition

In order to better understand the concept of master data it is useful to observe the concept from a broader perspective. McGilvray (2008, p. 39-44) divides different data categories as follows: master data, transactional data, reference data, metadata, historical data, and temporary data. Also configuration data can be represented as its own category (Kogent Learning Solutions 2011). These are groupings of data with common characteristics or features. Historical data and temporary data are additional categories that impact how systems and databases are designed and data are used. (McGilvray 2008, p. 39-44.) Different data categories are defined in the table 2.2.

Table 2.2. *Definitions of data categories according to McGilvray (2008) and Kogent Learning Solutions (2011).*

Data category	Definition	Examples
Master data	Describes the people, places, and things that are involved in an organization's business.	people (e.g., customers, employees, vendors, suppliers), places (e.g., locations, sales territories, offices), and things (e.g., accounts, products, assets, document sets)
Transactional data	Describes an internal or external event or transaction that takes place as an organization conducts its business.	sales orders, invoices, purchase orders, shipping documents, passport applications, credit card payments, and insurance claims
Reference data	Sets of values or classification schemas that are referred to by systems, applications, data stores, processes, and reports, as well as by transactional and master records.	lists of valid values, code lists, status codes, state abbreviations, demographic fields, flags, product types, gender, chart of accounts, and product hierarchy
Metadata	("data about data") labels, describes, or characterizes other data and make it easier to retrieve, interpret, or use information.	field names, length, type, lineage, database table layouts, field definitions, report names, application screen names, data quality statistics, the parties accountable for data quality for a particular

		field, timestamp, creator, create date, and update date
Configuration data	Configuration data is linked to the application processing logic and stored in configuration tables.	SAP ERP system configures the way of processing the data. Users can customize the configuration table according to the requirement of their company.
Additional data category	**Definition**	**Examples**
Historical data	Contains significant facts, as of a certain point in time, that should not be altered except to correct an error. It is important to security and compliance. Operational systems can also contain history tables for reporting or analysis purposes.	point-in-time reports, database snapshots, and version information.
Temporary data	Temporary data is kept in memory to speed up processing. It is not viewed by humans and is used for technical purposes.	a copy of a table that is created during a processing session to speed up lookups

Master data is the core data and the main component of most information systems. It is very essential to execute the business logic. (Kogent Learning Solutions 2011.) Characteristic for master data is that it is created once, used many times and does not change frequently (Knolmayer and Röthlin 2006, p. 362). For example, if an organization maintains a customer master data containing name, address, and other details of its customers, the chance of modification of these customer details is very small (Kogent Learning Solutions 2011). In ERP systems master data is usually used across several application modules in the transactions that are registered (Haug et al. 2009, p. 1055).

Master data is directly related to other data categories. For example, many business transactions rely on the availability of master data, representing, for instance, customers, suppliers, or products (Knolmayer and Röthlin 2006, p. 362). Figure 2.3. represents the relationships between the main data categories. The figure indicates that some reference data is required to create master data and that master data is required to transactional data creation. Sometimes reference data is also directly needed to transactional data

creation. In addition, metadata is needed to better use and understand all other data categories. (McGilvray 2008, p. 43.)

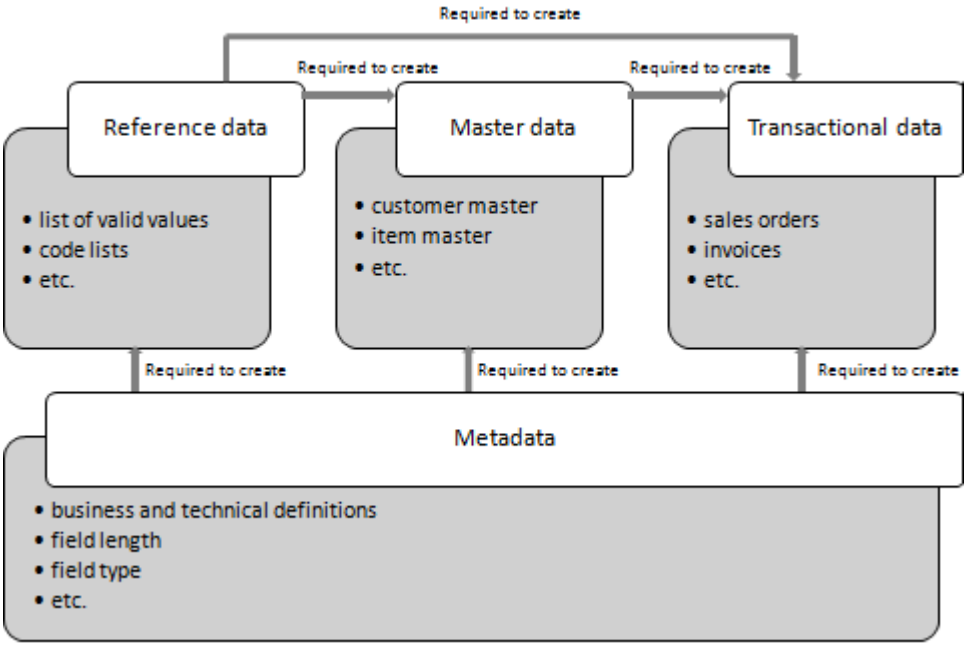

Figure 2.3. *Relationships between data categories (McGilvray 2008, p. 44).*

It is easy to understand from the figure above that also the care given to the reference data strongly impacts the quality of the master and transactional data. Reference data is the key to interoperability. The more it is managed and standardized, the more it increases the ability to share data across and outside of the organization. Errors in reference data have multiplying effects as the data continue to be passed on and used by other data. In addition, the quality of master data impacts transactional data, and the quality of metadata impacts all the other categories. For example, documenting definitions in the metadata improves quality because it transforms undocumented assumptions into documented and commonly agreed meanings so the data can be used consistently and correctly. (McGilvray 2008, p. 44.)

In conclusion, in order to improve the quality of the master data it is not enough to understand the concept of the master data itself but it is also essential to understand its relationship to the other data categories. Thus, understanding the different data categories and relationships and dependency between them is highly useful for managing the data because certain data may be treated differently based on its classification (McGilvray 2008, p. 39-44). In addition, from data quality perspective it is advantageous to understand the relationships between different data categories. For

example, since transactions use master data, if the master data is not correct, the transactions do not fulfill their intended purpose (Haug & Arlbjørn 2011, p. 289).

2.2.2 Sales & distribution master data

SAP ERP system is introduced in more detailed level later in the chapter three. However, at this point it should be understood that SAP ERP is a modular system. And more important each of these modules maintains master data (Kogent Learning Solutions 2011). Sales and distribution (SD) master data refers to the data which is an essential part of the SD module in SAP ERP system.

In SAP terms master data refers for example to company codes, cost centers, the collections of the products one sell, the materials one use to manufacture the products, the bills of materials with the components for each product, and the list of one's company's customers and vendors, etc. (Chudy & Castedo 2011, p. 41; Kogent Learning Solutions 2011). Each of these entities in the master data is composed of attributes, hierarchies, and tables (Kogent Learning Solutions 2011). In addition, master data in SAP refers to the rules governing the relationship with the business partners such as customers and vendors, detailing prices, discounts, terms, etc. (Chudy & Castedo 2011, p. 41).

The SD master data refers to a centralized data object that stores information related to the sale of products and services of an organization in the SD module of SAP ERP system. It reduces data redundancy and makes more efficient the transactions related to the customers, business partners, and the goods and services of an organization. Consequently, the SD master data is basically the data about the customers, materials, and the conditions on which sales depend. Therefore, the SD related master data can be further divided into three subcategories depending on the source of information: customer master data, material master data, and conditional master data. (Kogent Learning Solutions 2011.) The figure 2.4. illustrates the SD related master data.

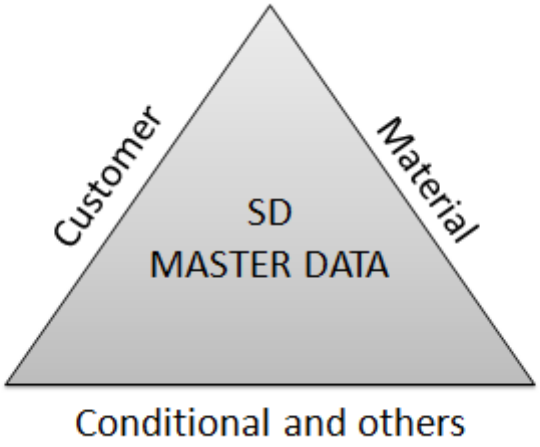

Figure 2.4. *Sales and distribution related master data.*

The customer related master data comprise a collection of data records related to customers, sales areas, and other transaction related units (Kogent Learning Solutions 2011). It includes all data necessary for processing orders, deliveries, invoices, and customer payments (SAP 2013a). This type of data aids in executing all transactions, delivery, and payment related functions in the SD module of SAP ERP system. The data related to all the customers is stored in the customer master data. (Kogent Learning Solutions 2011.)

The customer master data in SAP is grouped into three categories: general data, sales area data, and company code data. (SAP 2013a). The figure 2.5. graphically represents these three categories. The general data is a set of data independent of any organizational units of SAP ERP system. It is relevant for sales and distribution and for accounting. (SAP 2013a.) It is the central data set for a customer and it is valid for all organizational units within a client (Kogent Learning Solutions 2011). The sales area data is a set of data related specifically to the sales and distribution. It is valid for the respective sales area. (SAP 2013a.) Whereas, the company code data is a set of data used by the accounting department of an organization and it is valid for the respective company code (Kogent Learning Solutions 2011; SAP 2013a).

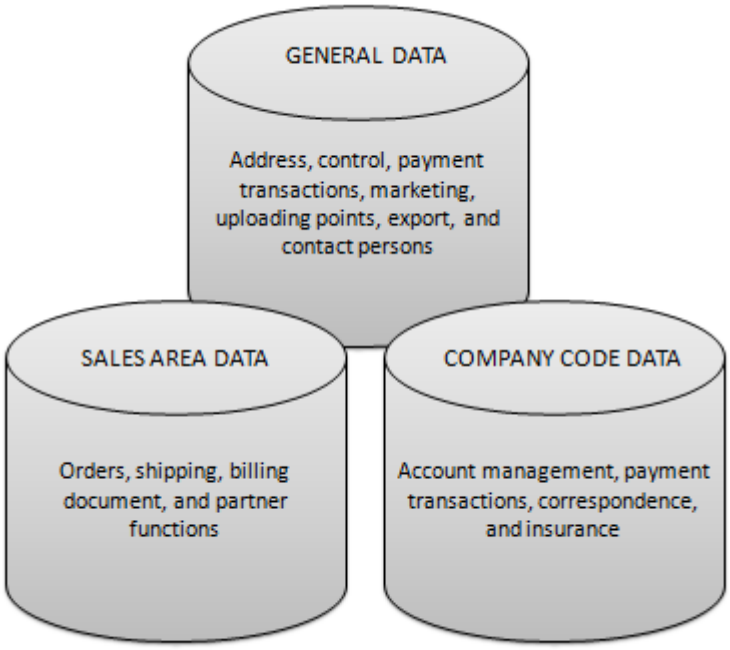

Figure 2.5. *The categories of the customer master data in SAP (adapted from Kogent Learning Solutions 2011).*

The data set related to material master data comprise all the data records related to the products and services offered by the organization (Kogent Learning Solutions 2011). Within the material master data, one can maintain detailed information about a certain material, e.g. description, weight, tax relevance (SAP 2013a). Because often many different departments, such as purchasing, sales, warehousing, or accounting, make use of material master data, the data is divided into several categories: basic data, sales data, purchasing data, accounting data, costing data, warehouse management data, etc. (Kogent Learning solutions 2011; SAP 2013a).

The basic data is a collection of general data of the materials related to all the organizational units of SAP system (Kogent Learning Solutions 2011). It is maintained independently of the organizational units (SAP 2013a). The sales and distribution data is related to the SD module. Whereas, the purchasing data is related to the materials management (MM) module. The accounting data is a collection of data related to the finance (FI) module. The costing data is used for cost determination and the warehouse management data used for warehouse management. (Kogent Learning Solutions 2011.) The relevant material master data for sales and distribution processes are especially basic data and sales data (SAP 2013a). The figure 2.6. graphically illustrates the different categories of the material master data.

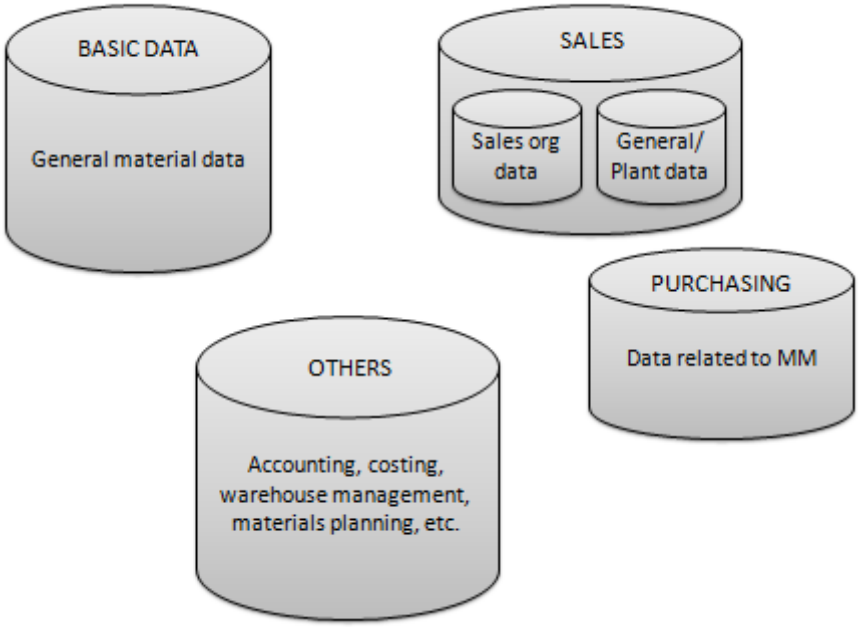

Figure 2.6. *The categories of the material master data in SAP (adapted from Kogent Learning Solutions 2011).*

Other SD related master data includes for example condition master data and customer-material information data. The condition master data comprise the business logic or configuration settings. It is determined using different condition techniques. (Kogent Learning Solutions 2011.) The condition master data includes master data for prices, surcharges and discounts, freights, and taxes. The different condition master records can be dependent on various characteristics, e.g. a material price could be customer specific or a discount dependent on the customer and the material pricing group. (SAP 2013a.) The figure 2.7. illustrates the condition master data in SAP.

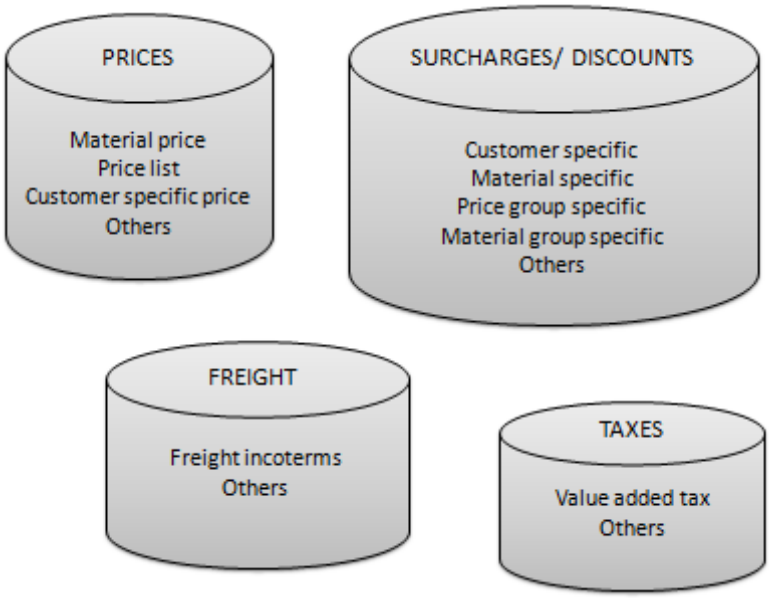

Figure 2.7. *The condition master data in SAP (adapted from SAP 2013a).*

The customer-material info data is related to the material and the customer who buys the material. Customers may have their own specifications for the materials which they are buying. These specifications can be stored as customer-material info data. (Kogent Learning Solutions 2011.)

The master data in SAP ERP system and especially in the SD module is the foundation on which transactions are executed (Chudy & Castedo 2011, p. 41). Thus, master data forms the basis of the entire data processing activities in SD. For example, when a customer places an order to purchase some goods or services, the customer's information is retrieved from the SD master data, which is used throughout the process of selling goods and services to that customer. In addition, the SD module retrieves information about the goods and services sold from master data. The SD master data also stores information about the sales agreements made between the organization and the customer. (Kogent Learning Solutions 2011.)

2.2.3 Effects of poor master data quality

Business data is the key business asset hence it needs to be actively managed to ensure it is accurate, current and fits for purpose (Marsh 2005, p. 106). If the quality of the key business asset is poor, it is doubtful how the business can operate accordingly. Thus, at the end of the day the consequences of the poor data quality can be huge. Madnick et al. (2004) also emphasize the fact that the data is the foundation for operational, tactical and strategic decisions of the companies which means that the high-quality data is crucial to a company’s success.

There are several data quality related surveys made by industry experts like Gartner Group, Price Waterhouse Coopers, and The Data Warehousing institute. The surveys claim to identify issues with data quality and data quality management in organizations. (March 2005; Haug & Arlbjørn 2011.) March (2005, p. 106-107) summarizes some of the results from these surveys as follows:

- 88 per cent of all data integration projects either fail completely or significantly overrun their budgets
- 75 per cent of organizations have identified costs stemming from dirty data
- 33 per cent of organizations have delayed or cancelled new IT systems because of poor data
- Business intelligence (BI) projects often fail due to dirty data, so it is imperative that BI-based business decisions are based on clean data
- Organizations typically overestimate the quality of their data and underestimate the cost of errors

The findings from these surveys clearly indicate the importance of understanding the causes for poor data quality (Haug & Arlbjørn 2011, p. 295).

Figure 2.8. represents the impacts caused by the data quality problems. Redman (2008, p. 46) divides the impacts in three levels: operations, tactics, and strategy.

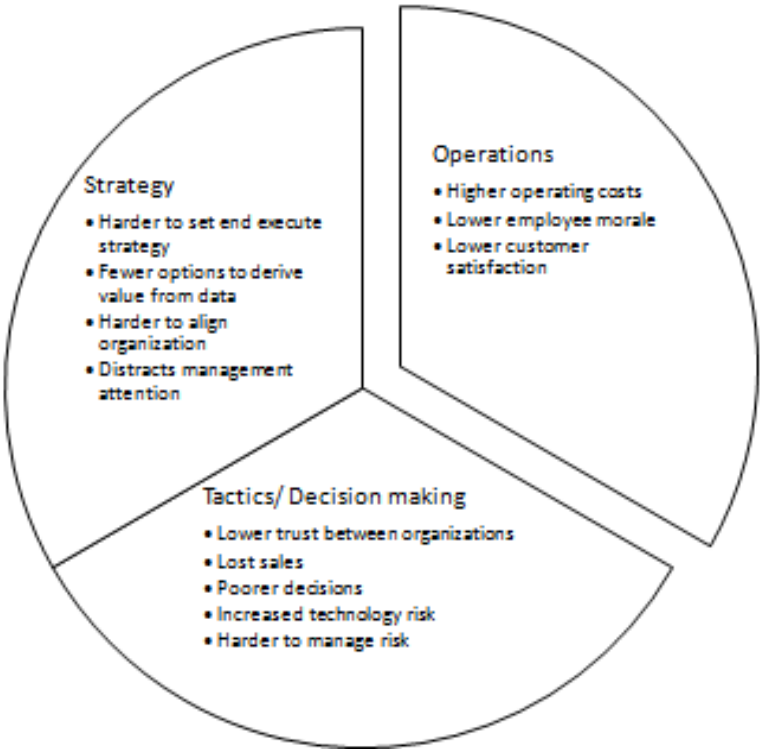

Figure 2.8. *The impacts caused by data quality problems by Redman (adapted from Redman 2008).*

From the operational perspective poor data quality increases the cost in operations. It also impacts negatively on employee morale and customer satisfaction. (Redman 2008, p. 46.) Also Marsh (2005, p. 106) argues that dirty data has a direct effect on stocking levels, sales orders, customer perceptions, loyalty and profitability. Redman (2008, p. 48) emphasizes the costs associated with customer dissatisfaction are important as no one really knows for example the lost opportunity costs of angering a customer with poor quality data. According to Olsen (2003, p. 14) customers that are lost because of poor quality data, e.g. continuous incorrect deliveries, flawed invoices, or other aspects of poor service, represent a large cost to the enterprises.

As the majority of the data in the sales or purchase orders is usually derived from the master data, the errors in the master data are often the source of the problems. For example, the issues such as incorrect delivery addresses, incorrect material numbers, and wrong prices are all based on the errors in the master data.

As an example, customers usually expect that their names and addresses are correct, the deliveries are handled properly and they are billed correctly but simple errors in the data get in the way (Redman 1998. p. 80). If considering the process of completing a customer order when all the data is correct the transaction is completed quickly and at low cost. But if any data, e.g. customer address, is missing or incorrect it has a direct effect on the process. (Redman 2008, p. 46-47) This impacts on the customer satisfaction as it takes longer to deliver the product to the customer and it can also frustrate the employees as they need to correct the errors and damages caused by faulted data.

In addition, data errors like wrong item numbers, wrong amounts, incorrect delivery addresses, etc. can cause multiple issues with normal business transactions. For example, when the wrong items are shipped and then returned, some costs occur. (Olsen 2003, p. 13.) These costs are called transaction rework costs. According to Olsen (2003, p. 13) some organization have entire departments that handle customer complaints on mishandled orders and shipments.

As another example, whenever the wrong parts or incorrect quantity of parts are delivered to the production line, there is either a stoppage of work or an oversupply that needs to be stored somewhere (Olsen 2003, p. 14). Thus, poor data quality increases operational cost because time and other resources are spent detecting and correcting errors caused (Redman 1998, p. 80).

Apart from the operations the impacts of the poor data quality extends far beyond them. Redman (2008, p. 46) concludes the impacts of data quality related to tactics as lower trust between (internal) organizations, lost sales, poorer decisions, increased technology risk, and harder to manage risk.

According to Redman (2008, p. 48) poor data is the main reason for the failures of new computer system implementations. Outstanding sums of money are spent in order to scrap or retrofit the new systems when the implementation project fails. Also Olsen (2003, p. 13) emphasizes that some of the major problems in system implementations are data errors and issues that block successful implementation. He states (Olsen 2003, p.13) that more than half of the times poor quality data increases the time and cost to implement data reuse projects by vast amounts.

Master data is the key data in any system implementation projects. It is essential to migrate the master data properly from the legacy systems to the target system as it act as a basis also for the transactional data. Therefore, errors in the master data can cause severe problems for the system implementations.

In addition, considering the decisions at any level it is said that decisions are not better than the data on which they are based (Redman 2008, p. 49). As Marsh (2005, p. 106) concludes if the data is incorrect, irrelevant, incomplete, ambiguous or inconsistent it can distort the true picture and lead to poor and costly business decisions. As the data is a critical input for almost all decisions it is impossible to make reliable and good decisions if the data quality is bad. Apart from the impacts on decision-making at the end of the day poor data quality can also lower organization's performance. (Haug & Arlbjørn 2011.)

On the other hand, it is clear that all decisions include some amount of uncertainty. However, decisions based on the most relevant, complete accurate and timely data have a better chance of advancing the enterprise's goals. (Redman 1998, p. 81.) Olsen (2003, p. 13) also stress that if the data sources are plagued by quality problems it generally requires manual massaging of the information before it can be used for decision making. However, if it is necessary to cleanse the data before decision making one can never be sure if the data is entirely correct after the clean up (Olsen 2003, p.13).

Poor-quality data affect strategy in a myriad of ways. If the data is bad and there is no technology to exploit that, it is hard to create strategies on the strength of the data. For example, if one cannot assemble a complete view of its customers, it cannot utilize the opportunities to cross-sell, uncover new product niches, or craft strategies based on its most profitable customers. (Redman 2008, p. 49.) Thus, the consequences that stem directly from operational and tactical issues, like poor processes and ineffective decision making, are quite clear (Redman 1998, p. 81). Redman (2008, p. 46) lists the impacts of data quality related to strategy as follows: harder to set and execute strategy, fewer options to derive value from data, harder to align organization, and distraction for management attention.

Considering especially master data quality, companies typically incur costs from two sides. Firstly, the costs raise when master data is cleansed to ensure high master data quality. Secondly, the costs incur due to the data that is not cleansed as poor master data quality might lead to faulty managerial decision-making, faulty operations and the other impacts as well. (Haug et al. 2011, p. 170.) Therefore, poor master data quality impacts on all the levels listed earlier: operations, tactics, and strategy.

As it was earlier discussed transactions use master data. If the master data is not correct, the transactions do not fulfill their intended purpose. Thus, errors in master data can have major costs. For example, if the address of a customer is wrong the orders and invoices may be sent to the wrong address, if the price of a product is wrong the product may be sold below the correct price, or if a debtor account number is wrong an invoice might not be paid correctly. Therefore, even a small amount of incorrect master data can absorb a great part of the revenue of a company. (Haug & Arlbjørn 2011, p. 289.) In addition, Chudy and Castedo (2011, p. 41) state that master data might very well be the one piece to SAP ERP implementation that can make the difference between getting the expected results or not.

All in all, data quality and especially master data quality is absolutely required for an ERP system to function properly. Because of the integrated nature of ERP, if someone enters the wrong data the mistake can have a negative domino effect throughout the entire enterprise. (Umble et al. 2002, p. 246.)

3. DATA MIGRATION AS A PART OF ERP IMPLEMENTATION

McGlivray (2008, p. 6) claims that one reason why data and information quality is getting more attention nowadays is the need for business to see data and information brought together in new ways. Among other things examples include the requirement to have data available for decision support through business intelligence and data warehousing, to streamline business processes and information through ERP system, and to deal with the high rate of mergers and acquisitions. All these need data integration which means bringing together data from two or more different sources and combining them to get better use of the data. (McGilvray, p. 6.)

Information systems are considered as key enablers for most business activities such as operations and decision making (Knolmayer and Röthlin 2006, p. 362). Today companies are investing millions to implement the next generation of information technologies, replacing department-sized legacy systems with ERP systems (Redman 2008, p. 48). Considered as the central component of today's information system landscape ERP systems provide organizations with large application functionality, supporting a major part of all business activities (Knolmayer and Röthlin 2006, p. 362).

Firstly, this chapter introduces the widely used ERP system SAP and then discusses about SAP ERP implementation projects. Secondly, this chapter covers the data migration process which is significant part of ERP implementations.

3.1 ERP system and implementation project

ERP systems are comprehensive packaged software solutions which integrate the complete range of business processes and functions in order to present a holistic view of the business from a single information and IT architecture (Klaus et al 2000, p. 141). Their design solves the problem of fragmented information in large organizations (Davenport 1998, p. 123). In addition, they bring together various types of data from different business activities such as marketing, sales, production, human resources, or finance in a consistent model of the business (Davenport 1998, p. 121; Knolmayer and Röthlin 2006, p. 362).

This chapter introduces SAP ERP system which was the target system in the empirical research described in the chapter four. In addition, this chapter describes SAP system implementation project.

3.1.1 SAP ERP system

SAP applications were developed and introduced by the German company called SAP AG (Kogent Learning Solutions 2011). SAP ERP has been for years one of the dominating and best-known ERP systems (Lau 2005; Kurbel 2013, p. 127). It is often used as an essential part of enterprise-wide information systems (Lau 2005). However, the name of the ERP system has changed many times over the years. Many ERP users know SAP ERP as R/3 which was still the official name at the start of the new millennium. Even though many companies still work with R/3 programs in one way or another, the name used in the marketing and official documentation is known as a component of SAP Business Suite called SAP ERP. (Kurbel 2013, p. 127.)

SAP Business Suite is a comprehensive package of application systems (Kurbel 2013, p. 127). Apart from SAP ERP, SAP business suite includes components such as SAP CRM (customer relationship management), SAP SRM (supplier relationship management), SAP SCM (supply chain management), SAP PLM (product life cycle management) (Chudy & Castedo 2011, p. 18-20; Kurbel 2013, p. 127). However, the largest component of the business suite is SAP ERP. (Kurbel 2013, p. 127.)

Since 2004 SAP has used the technology platform called SAP NetWeaver as the basis for the application systems. SAP ERP system contains the core component known as SAP ECC (ERP Central Component) and some additional components running on the NetWeaver platform. On the program level the core of the system is still largely based on the earlier R/3 code (Kurbel 2013, p. 128.)

SAP ERP is used to execute the day-to-day operations of a company. A company can for example plan and execute the steps involved in its supply chain, from the sales forecasts, master production scheduling, production planning, production scheduling, purchasing, inventory management, and warehousing to the logistics chain that is involved in shipping, transportation planning, and yard management. (Chudy & Castedo 2011, p. 18.)

Kurbel (2013, p. 127) and Padhi (2013) states that SAP ERP covers the following major elements:

- SAP ERP Financials
- SAP ERP Human capital management
- SAP ERP Operations

- SAP ERP Corporate services

SAP ERP Financials covers external and internal accounting (Kurbel 2013, p. 127). It is a comprehensive integrated financial management software that empowers finance and costing departments. Human capital management is an integrated and complete solution for human resource processes. (Padhi 2013.) It basically covers the personnel management (Kurbel 2013, p. 127). SAP ERP Operations helps achieve operational excellence in key business areas such as procurement, logistics, product development, manufacturing, sales and service. Whereas, Corporate services application increases efficiency by providing solutions to corporate service functions and supporting administrative processes. (Kurbel 2013, p. 127; Padhi 2013.)

SAP ERP application is divided into functional pieces called modules, such as sales and distribution (SD), materials management (MM), financial accounting (FI), controlling (CO), production planning (PP), etc. (Chudy & Castedo 2011, p. 18). The various modules are used to handle all business activities of an organization, such as recording the payment of invoices, controlling financial accounts, and managing production resources. (Kogent Learning Solutions 2011.) Chudy and Castedo (2011, p. 18) list the modules that are included in SAP ERP as follows:

- Financial accounting (FI)
- Controlling (CO)
- Project system (PS)
- Sales and distribution (SD)
- Materials management (MM)
- Logistics execution (LE)
- Quality management (QM)
- Plant maintenance (PM)
- Production planning and control (PP)
- Human resources (HR)

This thesis concentrates on the SD module. With the help of integration with other modules such as materials management and logistics execution the SD module covers the entire order-to-cash process in a company (Chudy & Castedo 2011, p. 35). It is one of the oldest SAP modules that serve as an integral part of SAP ERP system. The module basically helps to manage the sales and distribution activities in an organization. (Kogent Learning Solutions 2011.)

The sales and distribution in a company covers the entire chain of processes from customer inquiry and sales orders to the delivery of products to the customer destination of choice through billing and payment collection (Chudy & Castedo 2011, p. 30). Thus,

the basic functionalities of the SD module in SAP ERP system are fulfilling customer requirements, shipping or delivering goods to customers (delivery process), and receiving payment for goods or services. In addition, using the SD module a company can for example input the customer sales price for a product, check for open sales orders, forecast future requirements, set price quotations for a product and generate sales orders and bills for customers, etc. (Kogent Learning Solutions 2011.)

The master data is the core data in the information systems and it is very essential for the business logic execution (Kogent Learning Solutions 2011). Chudy and Castedo (2011, p. 41) also emphasize that the master data in SAP ERP system and specifically in SD module is the foundation on which transactions are executed. Master data was already introduced in the chapter two. It was discussed that the SD module deals mainly with the customer and material master data and the conditional master data.

Like other ERP packages, SAP is also a complex system. ERP systems are required to meet various business requirements as well as country-specific requirements. It can be concluded that no ERP can cover all of organizations' business requirements. However, SAP covers the standard functionality of almost all countries. (Padhi 2013.) In addition, if the system needs to meet the business requirements, also the business needs to meet the system requirements. The next two chapters concentrate on introducing the ERP implementation and especially SAP ERP implementation project.

3.1.2 ERP implementation

To understand better the ERP implementation project it is useful to introduce first the ERP system's life cycle. According to Souza and Zwicker (2009, p. 1431) the ERP system's life cycle describes the various stages through which a project of introducing an ERP system in a company passes through. The ERP implementation can be seen as a part of the ERP systems life cycle (Lau 2005). The figure 3.1. describes this life cycle and where the actual ERP implementation sits in there. The figure also explains what is characteristic for each phase in the life cycle.

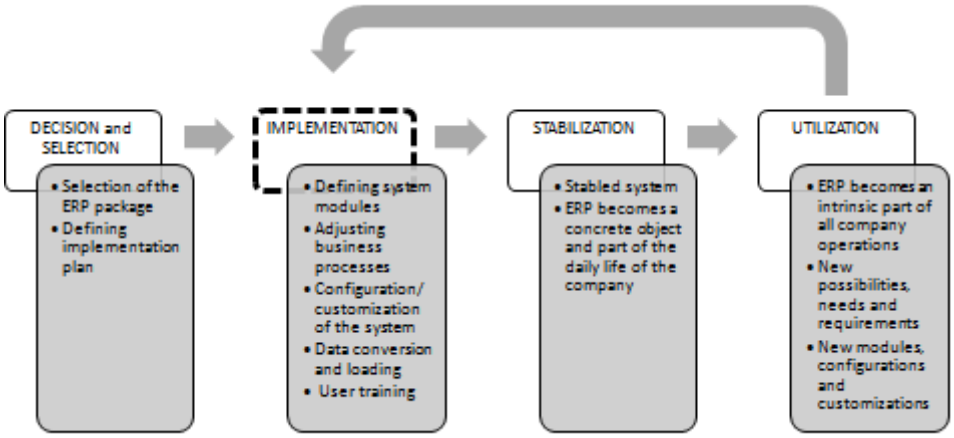

Figure 3.1. *The ERP system's life cycle (adapted from Lau 2005).*

The first stage in the lifecycle is the decision and selection where the company decides to implement an ERP system as the solution for its information systems requirements and chooses the supplier (Souza & Zwicker 2009, p. 1431). This stage also includes defining the implementation plan. The planning includes the activities such as establishment of the project's scope, defining the goals to be achieved and measuring methods for the verification of these goals, defining the project team and responsibilities, and choosing the implementation strategy. (Lau 2005.)

The ERP system implementation is the second stage of the introduced life cycle model. It can be defined as the process by which the system's modules are put into operation within a company (Souza & Zwicker 2009, p. 1431). This stage basically covers all the tasks that are carried out from the end of the implementation planning to the beginning of the situation when the ERP system becomes the company's official information system. Thus, the implementation stage includes the tasks such as configuration and customization of the system, the uploading of the data, hardware and software configuration, the adjustment of the business processes to the system, and user training. (Lau 2005.)

After the implementation stage the ERP system becomes a concrete object and part of the daily life of the company. However, all kinds of problems arise, such as operational problems, training deficiencies, and bugs in the software, which were not confronted at the implementation stage. Thus, there is considerable pressure to rapidly solve the problems. (Lau 2005.) In the stabilization stage these problems are solved and the ERP system is stabilized. This stage covers the first weeks after the beginning of an ERP system operation in the company (Souza & Zwicker 2009, p. 1431).

Lastly, at the utilization stage the system becomes a central part of all company day-by-day operations (Souza & Zwicker 2009, p. 1431). However, new possibilities, needs and

requirements may be noticed. Thus, this stage leads back to the implementation stage where the new modules, configurations and customizations can be implemented. (Lau 2005.)

So it is clear that the system implementation is part of the ERP system life cycle. When studying the implementation stage further, Kurbel (2013, p. 160) introduces three different strategies to implement an ERP system:

1. The company has their own implementation methodology.
2. The company works with the ERP vendor and applies the implementation methodology provided by the vendor.
3. The company works with an external consulting company and applies the implementation methodology recommended by the consulting company.

In the situation where the company has its own methodology it has usually been developed over the years and also applied for some other system implementations. The consulting companies may also have their own implementation methodologies based on the previous experiences. However, they can also apply the ERP vendor's recommended methodology. (Kurbel 2013, p. 160.)

Related to the ERP implementation strategies also different implementation approaches can be introduced. The figure 3.2. illustrates four ERP implementation approaches introduced by Lau (2005): phased in company level, phased in plant level, big-bang, and small-bang. The approaches are divided into four dimensions by the functional and the geographical dimensions. The functional dimension illustrates the number of modules that are implemented at the same time and the geographical dimension illustrates the number of localities or business units that begin to use the system at the same time (Lau 2005).

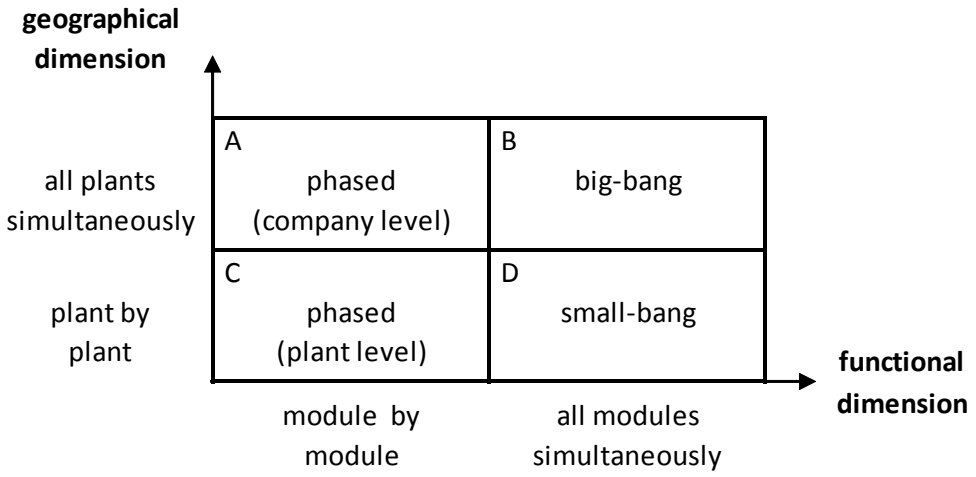

Figure 3.2. *The ERP implementation approach dimensions (adapted from Lau 2005).*

If the ERP system is implemented in a number of big stages each of which involves all modules being adopted simultaneously in one location or plant at the time, the used implementation approach is called a small-bang approach (region D in the figure 3.3.) (Lau 2005). On the other hand, implementing all modules of an ERP system in all locations or plants of the company simultaneously follows a big-bang approach (region B in the figure 3.3.) (Souza & Zwicker 2009, p. 1431).

The alternative choices for these are phased approaches which mean implementing an ERP system in sequential steps, comprising either more modules or more locations in one, until the system is completely installed (Souza & Zwicker 2009, p. 1431). For example, the implementation of one module at a time simultaneously in all locations characterizes a phased implementation at the company level (region A in the figure 3.3.). (Lau 2005)

It can be concluded that a company with a greater number of locations and implementing a greater number of modules runs greater risks than a company with fewer localities and fewer modules (Lau 2005). Thus, the big-bang approach can be seen more risky than the other approaches.

All in all, the ERP system implementation projects are considered difficult tasks that involve many organizational and technical issues (Lau 2005). For example, the fact that most companies installing ERP systems need to adapt or completely rework their processes and organization culture to fit the requirements of the system causes issues for the implementing organizations (Davenport 1998).

Kurbel (2013, p. 160) points out that in the implementation project there are number of factors that need to be considered. Thus, there is a great risk to forget some factor or choose the wrong approaches or solutions. Therefore, ERP system vendors and consulting companies usually recommend proven implementation methodologies which have already been successfully applied in other ERP projects. (Kurbel 2013, p. 160.)

3.1.3 SAP implementation

SAP is a complex system. It needs a different approach for its implementation than many other systems. (Lau 2005.) SAP brings its own culture which needs to be adapted within the existing organizational culture (Krumbholz & Maiden 2001, p. 186). Implementation of SAP also needs a fundamental change in the business processes (Mandal & Gunasekaran 2003, p. 275).

SAP has introduced an implementation methodology called accelerated SAP (ASAP). The main goal with the methodology is to speed up SAP implementation projects. (Lau 2005.) Basically, ASAP is a process model used for implementing SAP ERP systems.

However, it is not only a process model but it also includes computerized tools and other supporting features which help to deliver reliable results. (Kurbel 2013, p. 160.) The five key phases of the ASAP methodology are project preparation, business blueprint, realization, final preparation and go live & support (Lau 2005). The ASAP process is illustrated in the figure 3.3.

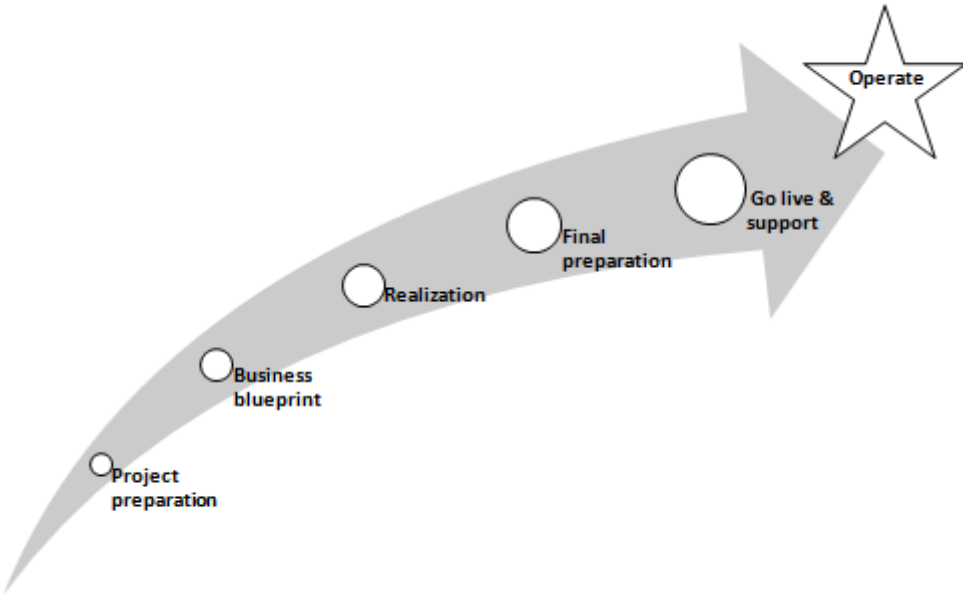

Figure 3.3. *The ASAP process phases for SAP system implementation (adapted from SAP 2014b).*

The ASAP process starts with the project preparation phase which provides initial planning and preparation of SAP project (Lau 2005; SAP 2014b). In this phase the primary focus areas, such as objectives, scope, plan and definition of project team, are identified and planned (Lau 2005).

The purpose of the next phase is to create the business blueprint. It is a detailed documentation of the results gathered during requirements workshops or meetings. (Lau 2005.) The goal in this phase is to achieve a common understanding of how the company intends to run SAP to support their business (SAP 2014b). The business blueprint phase allows the project team to clearly define their scope and only focus on SAP processes needed to run the organization business (Lau 2005).

In the realization phase all the business process requirements are implemented based on the business blueprint (SAP 2014b). The objectives are final implementation in the system, an overall test, and the release of the system for production operation (Lau 2005). The system configuration is an important part of this phase (SAP 2014b).

The purpose of the final preparation phase is to complete the final preparation, including testing, end user training, system management and cutover activities, to finalize the readiness to go live (Lau 2005; SAP 2014b). In this phase it is also important to resolve all critical open issues. When the final preparation phase is done successfully, the organization is ready to run the business in SAP system. (SAP 2014b)

In the fifth phase, go live & support, the goal is to go live with the new system (SAP 2014b). This means that the company starts to use the new system for its business operations and the system becomes a part of the daily life of the company. In addition, the long-term support needs to be organized and the system needs to be monitored and enhanced when necessary (Lau 2005). After the go live the company is operating and the business is running in the new system. At this point the system is the key element of the company and its operations.

When considering the ASAP implementation as a part of the ERP life cycle introduced earlier, it can be seen that the implementation project is not only a part of the implementation stage in the cycle. The actual project includes also tasks from the decision and selection stage and from the latter stages in the cycle as well.

3.2 Data migration

Morris (2012, p. 1) states that data migration projects have many forms. However, the classic form of data migration is where a new system is to be implemented and needs to be set up with a data from the legacy systems. (Morris 2012, p. 1.) Thus, data migration is also a crucial part of ERP implementation projects.

This chapter describes the data migration process in general and introduces the basic steps in the data migration. In addition, the chapter introduces how the data migration is usually handled in SAP ERP implementation. Finally, the challenges in the data migration process are discussed from different perspectives.

3.2.1 Data migration process in general

Haller et al. (2012, p.1) define data migration as a software-supported one-time process migrating data from a source system, which is supposed to be shut down, to a target system with a typically different data model. Morris (2012, p. 7) agrees stating that data migration is the one-off movement of the data from legacy systems to a new repository and it is intended to be one-way trip without return.

More precisely Morris (2012, p. 7) defines data migration as the selection, preparation, extraction, transformation and permanent movement of appropriate data that is of the right quality to the right place at the right time and decommissioning of legacy data

stores. He also emphasizes that the actual data migration does not involve the relocation of data centers or the regular movement of the data between business system and a data warehouse (Morris 2012, p. 7).

It has been stated that the pioneering approach for the data migration is a so called butterfly methodology (Haller 2009, p. 64). The objective of the butterfly methodology is to guide the migration of a mission-critical legacy system to a target system. The methodology is based on the assumption that the data of a legacy system is logically the most important part of the system. (Wu et al. 1997, p. 201-202.)

The butterfly methodology divides the data migration into five major steps: analysis, development of the data mappings, building up a sample data set in the target system, migration of the system components to the target system without any data, and step-by-step data migration (Hall 2009, p. 64). In addition, Wu et al. (1997, p. 203-204) have identified the sixth step known as cutover to the completed target system.

The earliest references to the butterfly methodology are from the 1990's (e.g. Wu et al. 1997). Thus, the methodology does not represent the timely approach for the data migration. Haller (2009) has introduced a model for a generic data migration architecture which provides a fresher approach to the subject. The generic data migration architecture is illustrated in the figure 3.4.

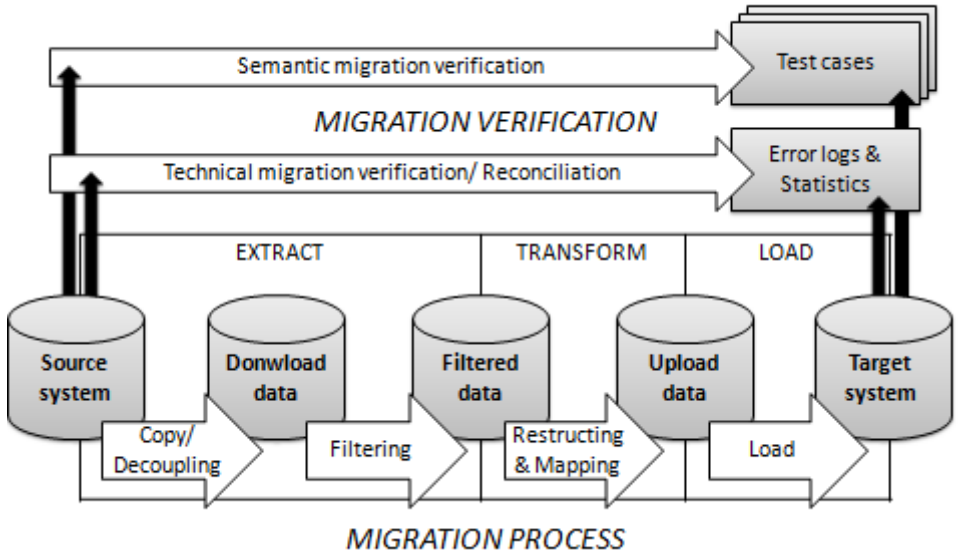

Figure 3.4. *The generic data migration architecture (adapted from Haller 2009, p. 65).*

The core of the generic data migration architecture is an ETL process (Haller 2009, p. 66). The term ETL stands for extract, transform and load (Morris 2012, p. 181). In the extraction step the first objective is to decouple the data. It means that the data from the

old system is copied from the old system to a different server. Thus, the data migration project cannot affect the business as usual in the legacy system. In addition, in the extraction step the data to be migrated is identified and everything else is filtered. (Haller 2009, p, 66.) For example, the customers or vendors which are not used anymore are filtered out of the data scope.

Next phase in the process is transformation. According to Haller (2009, p, 66) this means that if the data model in the legacy system does not correspond to the model in the target system the data must be restructured to fit to the target system. For example, if the payment term for 30 days net in the legacy system is indicated with the value "30" and the corresponding payment term in the target system is "Z300", the mapping and transformation of these values are executed in the transformation step. After the transformation the data is loaded into the target system (Haller 2009, p. 66).

After all the data has been migrated into the target system it is important to verify the results. In the model introduced by Haller (2009) this verification includes two techniques. The first one is for verification of the semantics of the data by choosing some test cases, e.g. important customers, which work as a sample objects. These objects are checked manually to ensure that all the fields and values are correct. The second technique is to reconcile that all the data is migrated rather than checking if the values are correct. (Haller 2009, p. 66.)

Morris (2012) studies the data migration process from a little bit different perspective. According to Morris (2012, p. 176-177) the design and execution of a successful data migration can be divided into following elements:

- Extraction including cleansing and preparation
- Transformation
- Testing
- Load
- Orchestration
- Fallback
- Legacy decommissioning

According to Morris (2012, p. 176) extraction, cleansing and preparation is nearly always the client's responsibility. However, the responsibility for transformation is split between client and supplier (Morris 2012, p. 176-177). The step which was not included in Haller's approach is testing. Morris states (2012, p. 176-177) that the testing is also a joint responsibility for supplier and client. Usually the supplier will perform unit testing, participate in integration and end-to-end testing, and be a leader in mock load testing (Morris 2012, p. 176-177). The detailed load design is nearly always the supplier's responsibility. (Morris 2012, p. 176-177.)

The approach of Morris does not separate the verification and reconciliation as a different step. However, Morris raises a couple of other elements which were not mentioned in Haller's approach. These are orchestration, fallback and legacy decommissioning.

Orchestration means the organization and management of the wider business to support the migration and to mitigate the disturbance caused by an enterprise application migration. This is always the responsibility of the client. Fallback design is a key part of any data migration design. It is usually a joint supplier/client responsibility. Fallback means the steps required to get an enterprise back into the position it was before the data migration. Legacy decommissioning is also a part of a data migration process. It is always the client's responsibility. (Morris 2012, p. 176-177.)

Even though the approaches introduced by Haller and Morris are a little bit different, there is still a clear common factor. This is the centre of the process: ETL. Thus, it can be concluded that the ETL process is the key element for any data migration process.

3.2.2 Data migration in SAP implementation

As it was already mentioned earlier SAP is a complex system and it needs a different approach for its implementation than many other systems (Lau 2005). SAP has established the rapid data migration solution which provides a process framework and the tools, methods, and templates needed to execute complicated migration tasks quickly, easily, and cost-effectively. (SAP 2014a.) In general, SAP's data migration solution provides reusable tools for ongoing data governance, pre-defined content and rules, and SAP & partner services (SAP 2013b). The rapid data migration is a process to facilitate the migration of critical information about customers, products, services, etc. to SAP solutions. (SAP 2014a.) The architecture of the rapid data migration solutions is represented in the figure 3.5.

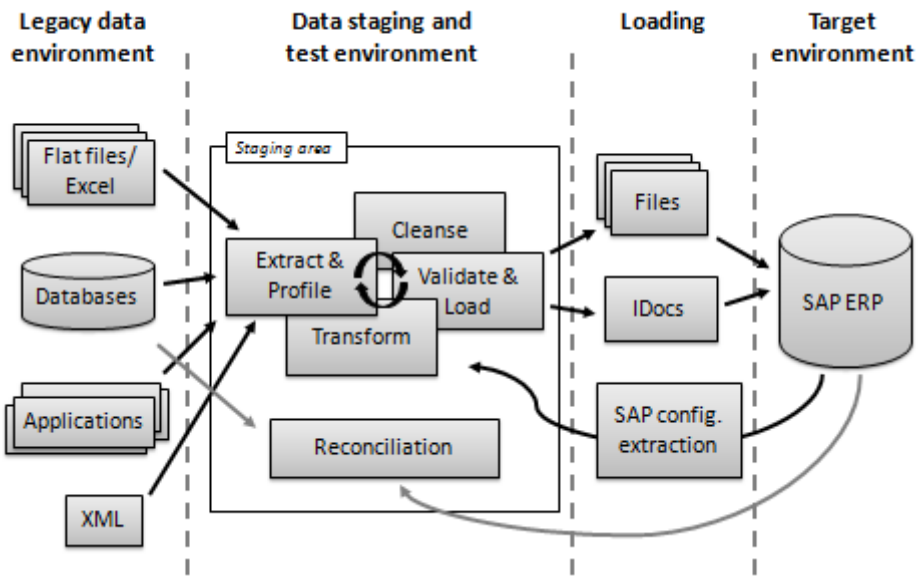

Figure 3.5. *SAP rapid data migration solution architecture (adapted from Sanjongco & Densborn 2013).*

Based on best practices SAP (2014a) introduces six distinct data migration activities which are included to the rapid data migration:

1. Analyze
2. Extract
3. Clean
4. Validate
5. Load
6. Reconcile

The source data destined for the target system may come from a number of sources: legacy databases, other ERP systems, Microsoft Excel spreadsheets, flat files, etc. (Sanjongco & Densborn 2013; SAP 2014a) The first step in the rapid data migration process is analyzing and profiling this source data in order to understand what one has and what kind of relationships are between different data components (SAP 2014a). This could include checking percentage of records that are complete as well as assessing how many follow a certain pattern (SAP 2013b). This initial analysis helps to assess the quality of the source data and clarifies the needs to cleanse and normalize the data before it can be loaded into the target system (SAP 2014a).

There are many practices and tools to assist in the analyzing activity. The tools can for example help to assess the status of the source data, compare it to the expectations of the target system, and identify areas where work is needed to help ensure a smooth migration (SAP 2014a).

The next phase is extracting the data from the sources. There are also tools which can act as a staging area where the data can be extracted from its source and prepare it for use in the target SAP software system (SAP 2014a). However, often the data is simply extracted and collected to Microsoft Excel spreadsheets and handled there. Sometimes the extraction can be also done before the analysis. Then the data analysis is executed after the data is extracted from the sources (SAP 2013b).

After the data is analyzed and extracted from the sources the next step is to cleanse the data. It is a crucial activity in the data migration. It helps to find errors in individual data objects, eliminate duplicate data entries, parse or merge data entries if necessary, and more important it helps to ensure the overall quality of the data destined for the target system (SAP 2014a). For example, in the legacy system one may have duplicate customer records with variant spellings such as “Boliden Ab”, “Boliden Mineral Ab”, as well as complete misspellings such as “Booliden Ab”. These duplicates and other data problems should be identified and cleansed in the data cleansing step. Sometimes it is also beneficial to execute the data cleansing already in the source system, which would make easier the migration itself (Haller et al. 2012, p. 172).

As a fourth step in the rapid data migration is the data validation against the business rules and the target system context (SAP 2013b). One important part of this step is mapping the source data fields to target data fields. This has to be done because the source system may identify the fields by one name and the target system by another name. (SAP 2014a.) The source system can also include fields which are not necessary or available in the target system or other way around. The data validation step also includes the transformation process which transforms the source data to meet the requirements and the data model of the target system (SAP 2013b).

In the data loading activity the data is loaded into the target system. There are different delivery formats to load the data into SAP including for example text flat files, XML (extensible markup language), BAPI (business application programming interface), IDoc (intermediate document), and ZTables (Sanjongco & Densborn 2013; SAP 2013b). The delivery format depends on SAP business object and the technology chosen for the data loading. The different technologies include for example LSMW (legacy system migration workbench), BAPI, and IDoc. (SAP 2013b.)

Finally, the last step in the rapid data migration process is reconciling the data. This means that the data loaded into the target system is compared with the source data in case of anomalies or failure data records (SAP 2014a). It is required to know what was supposed to be loaded versus what was loaded and to ensure that all the necessary data was loaded and is ready for use (SAP 2013b).

In all, the entire data migration process of assessing the data and validating it according to the business rules and SAP context lays the groundwork for ongoing data governance (SAP 2013b). Especially the analysis of the data, including assessing and profiling, is critical to any data migration project. Therefore, it is beneficial to conduct the data migration carefully, which also helps the future data management.

3.2.3 Barriers in data migration process

Applications tend to be long-living. They comprise thousands or millions lines of code. Small mistakes are guaranteed but also accepted and are corrected in later releases. In contrast, data migration comprises migration scripts, each possibly a few hundred lines of code. However, every small mistake can lead to a complete failure. For example, if a bank loses 0,1 percent of its customers during data migration, it is disastrous for the business. Thus, data migration should be executed completely correctly or it should be postponed. (Haller et al. 2012, p. 165) Therefore, it is also essential to recognize and overcome the possibly barriers or challenges in the data migration process.

Issues with data migration are costly. In addition, they can result lower profits and low productivity. Traditional approaches to data migration are often impacted by cost overruns, poor master data due to little to no data cleansing, and poor data quality in general. (SAP 2013b.) However, the success of a data migration project relies on several factors (Haller et al. 2012, p. 166).

Data migration process is all about moving the data from the legacy system to target system. Therefore, data quality is one of the most significant challenges to any data migration (Morris 2012, p. 7). Redman (2008, p. 48) also continues that poor data is the biggest reason for the high failure rate of new computer systems. Lack of quality data can make new system unusable and poor quality master data unable to support business processes, and return on investment is at stake if implementation is delayed (SAP 2013b).

Master data capture and processing are error-prone activities (Knolmayer and Röthlin 2006, p. 363). Errors made during master data capture and handling may increase through the system and also affect the quality of transactional data in the database of the ERP system (Xu et al. 2002). In the data migration the barrier is that both the data model and the representation of the data differ between the source and target application (Haller et al. 2012, p. 172). Thus, the data needs lots of modification before it can be migrated into the new system. Even where data is fine in its existing setting, it might not work in the new environment (Morris 2012, p. 7).

In addition, in the data migration it is essential to improve the quality of the data in order not to load bad data into the target system. However, according to Redman (2013,

p. 6) for most companies the barriers to improve the data quality are the facts that some managers refuse to admit that their data is not good enough. On the other hand, they might not even know how to fix poor quality data (Redman 2013, p. 6).

On the other hand, in the data migration lots of data is created into the target system. The best way to improve the data quality in long-term is to focus on getting new data right. (Redman 2013, p. 5) However, the barrier is that often the people creating the data have little understanding how others in the organization use the data. This is especially the case in the data migration projects. Thus, improving the data quality is about getting the creators of data to partner with the users, which however can be challenging (Redman 2013, p. 4).

Many data quality problems are rooted in metadata (Redman 2013, p. 6). As the metadata is required to create master data, the problems with the metadata have a significant impact on the quality of the master data. In addition, high quality metadata makes it easier for people to find the data they need, combine information, and draw the appropriate conclusions (Redman 2013, p. 6). However, in the data migration it is challenging to ensure the quality of the metadata.

Apart from the data quality issues the traditional barriers for data migration projects are lack of experience, unrealistic migration project plan, lack of confidence in the data, and the time external consultants are required to support the project. In addition, it is crucial to identify the right partner with the right expertise for the data migration project. (SAP 2013b.) Lack of experience refers to unskillful project team members. Haller et al. (2012, p. 166) emphasize that clearly defined and skillful teams and roles are necessary for a successful data migration. Apart from the planning also other project management issues cause barriers for the data migration (Haller 2009, p. 64). For example, it is hard to understand how difficult and time consuming the data migration can be (Eckerson 2002, p. 8).

Further possible barriers for the data migration are large data volume, a target database already containing data, and organizational challenges such as distributed project teams, which leads to poor communication (Matthes et al. 2011, p. 441). In addition, detecting failures, such as lost data, sufficient reconciliation and proper testing are barriers for any data migration projects (Haller 2009). Finally, the challenge of coordinating all data migration programs and invoking them in the correct order must be addressed in every data migration process (Matthes et al. 2011, p. 443).

When considering the data migration from SAP perspective, it has been stated that SAP brings its own culture which needs to be adapted within the existing organizational culture (Krumbholz & Maiden 2001, p. 186). Therefore, implementation of SAP, as

many other systems, needs a fundamental change in the business processes (Mandal & Gunasekaran 2003, p. 275). This can be a barrier for the data migration.

Furthermore, Matthes et al. (2011, p. 442-443) also demonstrates the data migration barriers as the risks involved in the data migration. These risks can be divided into three levels: business, IT management, and data migration. The three most relevant business risks are profitability, reputation and regulation. The IT management risks are data or information loss, target application stability, cutover aborts, extended downtime, project budget overruns, and delays in the project. Whereas the data migration risks are data completeness risk, data semantics risk, data corruption risk, stability risk, execution time risk, orchestration risk, dimensioning risk, interference risk, and target application parameterization risk. (Matthes et al. 2011, p. 442-443.) All of these risks can be seen as barriers for a data migration.

When considering the data migration as a part of ERP implementation, there can be many common challenges. Hawking et al. (2004) lists several barriers for the ERP implementation process:

- Lack of discipline
- Lack of change management
- Inadequate training
- Poor reporting procedures
- Inadequate process engineering
- Misplaced benefit ownership
- Inadequate internal staff
- Poor prioritization of resources
- Poor software functionality
- Inadequate ongoing support
- Poor business performance
- Under performed project team
- Poor application management
- Upgrades performed poorly
- Lack of management commitment
- Failure to include key personnel on the project
- Poor communication
- Poorly written or incomplete needs

These barriers can also be seen as challenges for the data migration process because if the ERP implementation fails most probably the data migration fails too. In fact, the listed barriers can be seen as challenges in all projects in general. This also makes sense because fundamentally the projects tend to have similar structures and processes.

On the other hand, when considering the data migration from data perspective, the general barriers or issues for data quality are also valid in the data migration. For example, rolled into one Haug and Arlbjørn (2011, p. 300), Umar et al. (1999), Xu et al. (2002), and Redman (2008) list the following factors influencing data quality in general:

- Lack of delegation of responsibilities for maintenance of master data
- Lack of rewards for ensuring valid master data
- Lack of master data control routines
- Lack of employee competencies
- Lack of user-friendliness of the software that are used to manage master data
- Lack of roles and responsibilities
- Lack of data quality owners
- Inefficient organizational procedures
- Lack of scheduling scenarios
- Lack of reward/reprimand system
- Neglecting administrative details
- Poor training
- Poor top management support
- Organization structure
- Ineffective change management
- Employee relations
- Data quality control
- People cannot find the data they need
- Incorrect data
- Too much data
- Data inconsistency across sources
- Poor data definition
- Data privacy and data security
- Organizational confusion

When comparing the ERP implementation barriers and the general data quality barriers and factors mentioned in the literature, several consistencies can be found. These factors can also be seen as barriers for data migration processes as the data is the key element in the data migration and on the other hand data migration is a part of ERP implementation.

Considering the data migration from a different perspective Morris (2012) has introduced four golden rules of data migration:

1. Data migration is a business not a technical issue

2. The business knows best
3. No organization needs, wants or will pay for perfect quality data
4. If you cannot count it, it does not count

These rules can be translated to the main barriers of the data migration:

1. Data migration is often seen only as a technical issue
2. The business does not understand that the final responsibility is in its own hands
3. The importance of data quality is not understood
4. Measuring the data quality is problematic

Data migration is often considered as a technical process even though fundamentally it is more a business process. Despite the fact that data migration normally occurs as a result of an IT project it is there to answer a business need. The business understands the business meaning and the value of its data and it is the value that must be preserved and enhanced in the data migration. (Morris 2012, p. 17.) The technology is there only to aid the business in the process.

It must be acknowledged that suppliers or consultants cannot know more about the business rules than the enterprise knows itself (Morris 2012, p. 23-24). However, the business often relies too much on the consultants and does not understand that they cannot do the migration alone. The business has the experience and expertise to make valid judgements about the quality and appropriateness of data items (Morris 2012, p. 23-24). The second golden rule covers the idea that the enterprise owns the data and the processes so it has the responsibility for decisions.

The third rule states that organizations are not willing to spend vast amounts of money for a perfect data quality. In addition, sometimes the organizations are not willing to spend money for data quality at all. However, as it was proven previously, data quality is a multidimensional concept and the quality depends on how the data is used. The data is high quality when it serves its purpose. The quality does not have to be perfect in order to serve the purpose. Nevertheless, it does not state that the data quality is irrelevant.

Well-managed projects have measurable deliverables and data migration project should not be an exception (Morris 2012, p. 27). Thus, the fourth rule indicates that the data quality and the progress in data migration process should be able to be measured. If the data quality in the target system should be “excellent”, it is challenging to evaluate whether this goal has been met. One needs hard facts about the data which can be measured (Morris 2012, p. 27). However, measuring and proving the data quality is often a major challenge.

In conclusion, data migration process has multiple barriers which can be seen as challenges for the process. They all have some effect on data quality as data is the key in the data migration. Even though data migration is basically moving the data from one system to another, in reality the challenges are much wider than just the actual migration of the data.

4. EMPIRICAL STUDY

The empirical study of this research was based on the theme-centered interviews. This chapter describes at a detailed level the empirical part of the research. Firstly, the chosen data collection and analysis techniques are explained and justified. After that this chapter clarifies how the empirical study was conducted.

4.1 Data collection and analysis

There are myriad methods to conduct the data collection and analysis of the research. The chosen methods depend highly on the nature of the research and on the chosen approaches and strategies for the research. Firstly, this chapter describes and justifies the data collection techniques used in this research in more detailed level. Secondly, the data analysis method is explained.

4.1.1 Data collection

Typical for a qualitative research is that data collection methods are qualitative in nature. These kinds of methods are different interviews and observation where the viewpoints and thoughts of the interviewees play important role in the research. (Hirsjärvi et al. 2005, p. 164.) Hirsjärvi et al. (2005, p. 135) describes that typical for case studies is the use of several data collection methods. However, interviews are said to be the most important sources of case study information. (Yin 2009, p 128) Also in this research the main data collection method was interviews.

There are different kinds of interview types, such as structured interview, semi-structured interview, and open-ended interview which is also known as unstructured interview (Aaltola & Valli 2007, p. 25-27; Saunders et al. 2009, p. 320). In addition, Aaltola and Valli (2007, p. 25-27) includes a theme-centered interview. In a structured interview the questions and the order of them are same for all interviewees. They are often multiple choice questions as the alternative answers are usually given. A semi-structured interview follows the same idea than structured interview but the alternative answers are not given beforehand. An open-ended interview resembles a normal discussion where the interviewer and the interviewee discuss about the certain topic but the themes are not same for all of the interviewees. (Aaltola & Valli 2007, p. 25-27.)

A theme-centered interview can be seen somewhere in the middle of open-ended interviews and semi-structured interviews. According to Aaltola and Valli (2007, p. 25-27) a theme-centered interview can be described as a discussion which commences on the researcher's initiative. It is often done with the researcher's terms where the discussion is concentrated on the subjects which the researcher sees important for the research topic. In other words, the certain themes for the interview are predetermined and the same themes are discussed with every interviewee but not necessarily in the same order. (Aaltola & Valli 2007, p. 25-27.)

In qualitative research the interviews are usually conversational in nature. The goal is to encourage interviewees to have the time and opportunity to reconstruct their own experiences and reality in their own words. (Yin 2011, p. 32.) Thus, the data collection cannot be founded on a questionnaire based data collection methods, such as structured interviews. Yin (2011, p. 102) also states that open-ended interviews are often preferred in case studies as the conversational style avoids steering interviewees as much as possible.

However, if the researcher has defined the study topic and started to articulate some key research questions some sort of protocol can help to guide the study and the data collection in a productive manner (Yin 2011, p. 102). Thus, the main data collection method which was chosen for this thesis was a theme-centered interview where the chosen themes work as protocols which guide the research. The relevant themes for the interviews were chosen based on the initial theory.

Nevertheless, the theme-centered interviews were not the only data collection methods used in this thesis. Based on the interviews and the initial analysis the results were thrown back for the interviewees in a form of a questionnaire. Thus, the simple structured multiple-choice questionnaire brought more valuable data for the study and the primary research problem. There are a range of different questionnaire techniques but in this thesis the used technique was an online questionnaire.

The term questionnaire refers to a technique of data collection in which each person is asked to respond to the same set of questions in a predetermined order (Vaus 2002). In addition, the term online questionnaire refers to self-administered internet-mediated questionnaire where the interviewer is not present (Saunders et al. 2009, p.362).

Tashakkori and Teddlie (2010) state that multiple methods are useful when they provide better opportunities to answer the research questions and where they allow to better evaluate the extent to which the research findings can be trusted and conclusions made from them. Saunders et al. (2009, p. 153) also continue that multiple methods can give a confidence that one is addressing the most important issues. Also in this thesis the multiple methods supported each other and brought new insights into the research.

4.1.2 Data analysis

Hirsjärvi et al. (2005, p. 164) state that characteristic for a qualitative research is that the analysis is inductive. This means that the starting point is not to test theories or hypotheses but to examine the research material in detailed and multifaceted level (Hirsjärvi et al. 2005, p. 164). However, as it was stated already in the introduction, in practice the research is likely to combine elements of both inductive and deductive approaches (Saunders et al. 2009, p. 490). Thus, the analysis mainly followed the inductive approach but it might have also some deductive characteristics. In any case the purpose in this thesis was to analyze the interview material versatilely and in depth.

The primary research data was qualitative in nature as it was collected via interviews. However, there is no standardized procedure for analyzing such data (Saunders et al. 2009, p. 490). In spite of that Saunders et al. (2009, p. 490) list three main types of procedures for analyzing qualitative data:

- Summarizing
- Categorization
- Structuring using narrative

Summarizing means compressing long statements into briefer statements where the main sense is rephrased in a few words (Kvale 1996 in Saunders 2009, p. 491). Thus, it involves condensing the big amounts of text into fewer words. By summarizing the data one understands the main themes which have emerged from the interviews and is able to identify clear relationships between them. (Saunders 2009, p. 492.)

Categorizing data includes two activities: developing the categories and organizing the data according to the categories. By doing this one will understand the relationships and is able to already draw conclusions from the data. (Saunders 2009, p. 492.)

Structuring the data using narrative means analyzing the data in its originally told form rather than by seeking to fragment it through categorizing or coding (Saunders 2009, p. 497). A narrative can be defined as an account of an experience that is told in a sequenced way including related events which are significant for the narrator and which convey meaning to the researcher (Coffey & Atkinson 1996 in Saunders 2009, p. 497).

In this thesis the main data analyzing procedures were summarizing and categorizing the data. Structuring the data using narrative did not really fit into the research objectives and it was noted that it would not have provided any extra insights into the analysis. In addition, after the summarization and classification the qualitative data could also be quantified. Especially by means of the additional questionnaire conducted after the interviews the qualitative data was quantified. By quantifying the data it was

possible to include a useful supplement to the principal means of analyzing the qualitative data (Saunders 2009, p. 497).

4.2 Conducting the study

In the beginning of the research project the initial theory was reviewed and the schedule for the project was created. According to the initial theory review the interview protocol was created. Eventually, the finalized interview protocol included ten themes: (1) data quality, (2) perfect data quality, (3) data quality problems, (4) data migration process, (5) problems in the data migration process, (6) data quality barriers in the data migration, (7) improving the data migration process, (8) improving the data quality in the data migration, (9) the lessons learned, and (10) prospects. The finalized interview protocol is described in the appendix 1.

The interviewees were carefully chosen from people who have knowledge and experience of the studied phenomenon: data migration in SAP implementation. The aim was also to have interviewees with different roles and backgrounds in order to get a broader perspective for the research. The potential interviewees were identified from the ongoing SAP implementation and migration projects. At first, the potential interviewees were contacted via e-mail and the suitable times for the interviews were agreed. No material was provided for the interviewees before the interviews.

Eventually, in total nine SAP consultants with different backgrounds were interviewed for the research. The one-to-one theme-centered interviews were held in December 2014. The interviews were held in English or in Finnish depending on the interviewees' native language which was Finnish for five of the interviewees. The rest were held in English.

The final group of interviewees is listed in the table 4.1. The table describes how the interviewees are divided into different groups according to their answers to the background questions. The interviewee ID is used later when referencing to the interviews in chapter five.

Table 4.1. *The interviewees of the research.*

ID	Current project role	Group	Relevant experience (years)
I1	Data team lead	Managing	16 - 20
I2	Solution area lead	Functional/ Managing	6 - 10
I3	Cutover manager	Managing	6 - 10
I4	Solution area lead/ FICO	Managing	11 - 15
I5	Data migration consultant	Technical	1 - 5
I6	Solution area lead	Functional/ Managing	16 - 20
I7	Data migration consultant	Technical	16 - 20
I8	Data migration consultant	Technical	11 - 15
I9	Data migration/ SAP SF consultant	Technical	1 - 5

During the interviews PowerPoint slides were used to give a short preface for the research itself, to clarify the research scope, to display the background questions, and to indicate the themes and the progress of the interview. It was especially emphasized that the research concentrates on the SD master data and data migration in SAP implementation.

The basic background questions, such as current project role, best suitable group, and relevant experience, were asked from all interviewees. The predetermined groups were managing consultant, functional consultant, and technical consultant. The managing consultant group represents people who are in a leading position and manages some team, process, solution area, etc. The functional consultants are the persons who understand the modules they are implementing and the setup and configuration options available. They are involved in the planning, designing and control of the system configuration. The group of technical consultant includes persons who do the hands on data migration job such as data loading, etc.

Three of the interviewees considered themselves as managing consultants and four as technical consultants according to their current project role. On the contrary, two interviewees stated that their role is both functional and managing. However, many of the interviewees had also years of experience in several different roles depending on the length of their career. Thus, all the interviewees had been in data migration related project roles at some point in their career. Nevertheless, it should be noted that the interviewees were not master data management experts but data migration and SAP system subject matter experts. Therefore, they had a lot of experience of SAP implementation projects and how the data should be managed during the projects.

All themes in the interview protocol were discussed with each interviewee but not necessarily in a same way and same stress. It highly varied how much the interviewees had to say about the different themes. Also the length of the interviews varied from 35 minutes to 90 minutes. All the interviews were recorded if the interviewee gave the permission and transcribed after the interview.

The interviews were usually transcribed almost immediately after the interview by listening the recording. At the same time the material were summarized into clearer and compressed format and initially categorized into different groups developed according to the results. After each interview the new results were compared to the other interviews and the categorization was adjusted after each round. Thus, the analysis was more or less incremental and continuous process.

After all the interviews were held and the categorization was done the additional questionnaire was created. The purpose was to create a simple questionnaire to support the primary research problem. Thus, the questionnaire included only one question where one had to select five most intrinsic methods to improve the master data quality in the data migration process of an ERP implementation. The question was a multiple-choice question which alternative choices were collected from the interviews. The questionnaire was sent back to the interviewees as an online questionnaire. They were given two weeks time to take the questionnaire. The simplified questionnaire is represented in the appendix 2.

After all the interviewees had answered to the questionnaire the results were analyzed. According to the questionnaire results the most important results could be recognized and pointed out from the initial list of the methods to improve master data quality in the data migration.

5. RESULTS

This chapter concludes the results from the interviews and from the additional questionnaire. The interview results are divided into four different groups according to the themes included in the interview protocol. The first subchapter discusses the results from the first three interview themes: data quality, perfect data quality, and data quality problems. The second subchapter includes the next three themes: data migration process, problems in the data migration process, and data quality barriers in the data migration. This chapter reflects the challenges in the data migration process for the data quality. On the contrary, the third subchapter concludes the results from the themes improving the data migration process, improving the data quality in the data migration, and the lessons learned. This chapter also concludes the results from the additional questionnaire which was conducted after the interviews. Finally, the last subchapter includes the last theme from the interviews: prospects. The interviews focused especially on the SD master data but in this chapter it is often referred to only as data.

5.1 Data quality and perfect data quality

The first themes in the interviews concentrated on the data quality in general. Commonly, data and data quality were seen as interesting subjects to discuss and they were stated to be very important for the businesses.

"Data is an asset, data is a business enabler." (I1.)

"Data is one of the most important things which needs to be in place because no business process runs without data." (I3.)

"All you do in the system is dependent on the quality of the data. The data quality is a key when you start an ERP system" (I7.)

In the interviews data quality was often referred to as correct, accurate, complete, reliable, cleansed, and up-to-date data. Thus, it can be stated that the multidimensional nature of data is well known. However, it was also noted that data quality is a relative concept. Thus, it is important to understand the purpose of data.

"Data quality is good when the data serves its purpose in SAP." (I4.)

"Good quality data works for the purpose of the business" (I8.)

In addition, it was said that data quality defines how a business will run both effectively and efficiently. Thus, there is a correlation between data quality and the effectiveness of a business. It was also agreed that data quality is definitely not overrated.

"Data quality has a direct impact on customers' perception and the efficiency of a business." (I1.)

When the discussion moved to perfect data quality, it was agreed that it is definitely something that should be tried to achieve. However, similar understanding was about the fact that it is almost impossible to reach in practice as it requires too much effort. The main reason for this is that it is very expensive to acquire the necessary tools and time-consuming to perform the continuous cleansing which is required to in order to have perfect data. Also the fact that nowadays the volumes of data tend to be huge makes it even harder to achieve perfect data quality.

"Perfect data has also a perfect price." (I6.)

"Perfect data quality is not possible but the data quality should be at the level so that it does not affect your business." (I7.)

The resources should be focused especially on the most important data objects from the business perspective. Often the most important data objects are customers, vendors and materials. One of the interviewees referred to these as the big three data items. It was also noted that especially SAP is very sensitive to data quality in order to work as it should. Therefore especially the system critical data should be in place.

"Without questions the big three data items should be in good condition." (I1.)

"The most important data objects should be taken care of really carefully." (I2.)

When talking about data quality problems it came clear the problems are usually quite similar. There are always duplicates, invalid data, and incorrect data to cleanse. Often the material data was claimed to be the most problematic data object to handle especially due to the huge volume. But also customer data was mentioned as a complicated object due to the different partner functions in SAP.

As a conclusion, there was a common understanding about the importance of the data quality within the interviewees. It was also agreed that perfect data quality is not a necessity but it is essential to concentrate on the most important data objects for the business. The discussion related to data quality problems varied depending on the interviewee. The discussion also started to spread to the following themes. Thus, the next chapter concentrates more on the data migration process.

5.2 Data quality barriers in the data migration process

The interview themes from four to six concentrated on the data migration process and problems in the data migration. In many interviews it came clear that SAP has proven methods how to move someone into a new system. The interviewees referred for example to the ASAP methodology and rapid data migration solution which are developed especially for SAP systems implementation. In general, the interviewees agreed that these are well-proven methods and they follow the good quality standards.

However, at this stage the interviews mainly concentrated on the problems in the data migration process. Especially the problems considered as data quality barriers in the process were discussed more thoroughly. The data quality barriers were discussed especially from the master data perspective.

According to the interviews the often confronted problem in the data migration is that the data migration process is seen only as a technical process. Another common error is that the data experts are included in the project only at the later steps and the role of the data team is underestimated. Thus, they miss the good opportunity during the initial preparation stage of highlighting the importance of data to the business and the end users.

"Issue with many organizations is that they concentrate initially on the technical solution which is also important but sometimes they make a mistake on boarding the people involved with the data only at the later steps of the project." (I1.)

Another problem is that often the importance of data quality is underestimated. Businesses also tend to think that the data quality is good enough in their legacy system. However, often the truth is the opposite. In reality, businesses do not usually know the status and quality of the data in their legacy system. The root cause for this can be that sometimes businesses don't have operating master data governance who maintains the quality of their master data in the legacy system. Thus, the requirement for improving data quality can come as a surprise.

"Often the quality of the data in the legacy system is misevaluated. One thinks that the migration happens in the wink of an eye" (I5.)

"The usual problem is the quality of the data in the legacy system. Especially, the SD data objects tend to be old and without maintenance for years." (I3.)

From a consultant perspective it is a whole different thing in presenting back the status of their data. One should have the tools to be able to analyze the data and present that back to the business. However, businesses do not usually want to spent money upfront for proper data tools. According to the interviews there are lots of good tools for taking a

feed from the data and being able to quickly analyze the data and push it back to a very easy-to-understand format. However, lots of companies don't want to spend money on those and the consultants are basically dependant on the tools provided by the business. The problem is that the data is not seen as an investment.

"By neglecting the data the costs can be very high. It is not a lot of money to be spent upfront to have proper tools to use." (I1.)

Change management is also important from the data quality perspective. It can also be a barrier for successful data migration process. People tend to fear the change. However, they need to understand the importance of the data quality and they really need to buy in the idea of that. Sometimes people also need to be convinced that the change is good.

"About 15-20 years ago around 80% of SAP projects failed because of two reasons: data and people. People are related to change management. Data is a well known factor why projects are failed or ran over budget. Really the failing occurs when the people fails to acknowledge that." (I1.)

"People are very sensitive with their data and they don't want it to be changed. However, in system development projects it is usual that the data structures are changed which frighten and confuse people." (I5.)

Another barrier for the data quality in the data migration is unclear project scope. If it is not clear which processes, applications and data objects should be migrated, it is impossible to end up with good quality data in the new system. In addition, if the project does not have enough resources or the project members have poor commitment to the project, it is very likely to be unsuccessful. Also poorly trained or unskillful project team is a barrier for the data quality. For example, if the systems are not familiar enough, successful data migration cannot be ensured.

"Data migration project related tasks are often only the third priority for the business and they put their business tasks before the project tasks" (I2.)

On the contrary, the data migration is often poorly planned and the schedule is underestimated. Related to the planning also roles, ownerships and responsibilities can be unclear within the project team.

"In every project I've seen that the data migration process and the schedule are underestimated." (I3.)

"Sometimes the roles and responsibilities are understood differently. They are written in the contracts and plans but not necessarily understood." (I4.)

Even if the project team was in place, another problem could be the communication. The communication can be problematic either within the project team or to the customer side. The reason for this can be ineffective communication procedures or the language, i.e. jargon, used in the communication. The customer does not necessarily understand the technical terms used in the project. The customer can also have very determined decisions which are not necessarily reasonable.

"Communication is always a challenge. You have to avoid speaking rocket science." (I7.)

"Sometimes the customer does not understand. No common language. The customer asks to implement strange things just to make things happen and be like in the legacy system." (I8.)

Sometimes also huge volumes of data can be a barrier for the data quality. Firstly, one might need tools and more processing power for the millions of records of material data. They are harder to handle and maintain during the migration process. Secondly, legacy systems usually also have some irrelevant data which is not used anymore. If this data is migrated, it lowers the quality of the data.

Also documentation and file handling can be a problem if it is not disciplined. Sometimes the files are just saved in some local drives or random folders. Consequently, the version history and the naming convention are not clear enough. This causes troubles for the transparency, validation and audition of the project. In addition, this causes problems especially in case of project team changes.

"If the project team changes, it is really hard to take over without clear instructions and disciplined documentations." (I2.)

Often the root cause for the barriers is that the legacy system and the target system are totally different. For example, the data models in the legacy system and the target system are usually not consistent. Thus, it is necessary to reconstruct and transform the data and map it to fit to the target system. Also the fact that sometimes the legacy system infrastructure and data models are not documented at all makes the mapping more difficult and causes a barrier for the data quality.

"The quality problems are related to the legacy system and the mapping. SAP has strict rules. Moving from not so strict environment to SAP is not an easy task." (I7.)

"Data modeling and mapping problem is common. The legacy system thinks differently than SAP. There are no one-to-one relations between all fields or even data objects." (I4.)

"For example, you have pricing in the legacy. In sap you may have pricing with discounts. You may have to translate the net price into problematic discounts, which is hard." (I7.)

According to one interviewee data migration should be business process translated to technical process. One has to make many decisions based on business reasons. In other words, the business values have to be translated into technical definitions. For the translation one needs clear rules. However, the definitions for the business processes and the rules for the translation are not always clear. Thus, it can be a data quality barrier in the data migration process.

Also the rules and criteria for the data quality are needed. The required data quality level cannot be entirely known if there are no clear criteria. Sometimes the business reasons for the data quality requirements can also be unclear. On the other hand, occasionally the whole data migration process and what it all requires is not understood properly.

"Customers rarely have clear criteria and required quality levels for the data. It is often thought that if they have an operating legacy system, everything is fine. It is just a migration." (I4.)

In one interview it was noted that data is often siloed for different persons and organizational units, which causes problems like duplicate records. Also the fact that people do not often understand the data or the purpose of it leads to data quality problems such as incorrect data.

"Often the data is siloed for different organizational units and people don't communicate and utilize that data." (I5.)

In addition, when considering especially master data people see it only as a static data. Even though master data does not usually change frequently it can still become outdated. Therefore it easily becomes incorrect or invalid if not maintained regularly.

"Master data is often thought as a static thing even though it is a process. It has a life cycle including creation, use, maintenance and deletion or archiving." (I6.)

Also technology can be a barrier for the data quality. For example, if the data uploading techniques are challenging, it can be a risk for the data quality. In addition, human error has also its role in the data migration process. There are lots of manual tasks in the data migration process and the data can also be handled manually. Thus, there is a big possibility for human errors. In addition, if everyone has access and possibility to handle the data, there is even greater risk for errors.

In complicated projects there are always surprising changes. However, if these changes are not communicated properly, it can cause problems for the data migration and data quality. The normal projects have also usually proper tests where the software and the processes are tested before the go live. However, often the data is not tested properly due to timing issues or just because the importance of the data is neglected.

"Projects have tight schedules where might be hurry to load some data for the tests but the real data cannot be tested properly." (I2.)

As a conclusion the table 5.1. lists all the data quality barriers in the data migration process which were raised in the interviews. Each of them affects negatively the master data quality in the data migration process.

Table 5.1. *The data quality barriers in the data migration process.*

The data quality barriers in the data migration process
Data migration is seen only as a technical process
Data experts are included in the project only at the later steps (not from the beginning)
The importance of data is underestimated
Businesses do not want to spent money upfront for proper data tools etc.
Businesses do not know the status and quality of the data in the legacy system
Inefficient management of change
Weak buy in for data quality from the business
The data migration project scope is not clear
The project does not have enough resources
Project people do not have commitment for the project
The data migration project is planned poorly
The data in the legacy system is poor quality
Data migration process and the schedule is underestimated
There is no communication within the project team or it is hard, e.g. difficult terms, jargon, etc.
Huge volumes of data
The role of the data team is underestimated
Documentation and file handling is not disciplined
Roles, ownerships and responsibilities are not clear within the project team
Not operating master data governance team for the legacy system
Difficult customers
Data conversion is hard from the legacy to the target
No clear data quality rules or criteria
The data is siloed for different people and units
Data is not seen as an investment
The requirement for improving data quality is a surprise
The fear of the change

Data models in the legacy system and the target system are not consistent
The legacy system infrastructure is not documented at all
It is hard to translate the business requirements to technical terms
Incorrect or unclear rules and definitions for business processes
Poorly trained or unskillful project team
The business reasons for the data requirements are unclear
Businesses don't understand their data
Master data is seen only as a static data
Data migration process is not understood
Data uploading techniques are challenging
Surprising changes in the project, system etc.
Poor technical tools
Human errors
Everyone has access and possibility to handle the data
The systems are not familiar enough for the project team
The data is not tested

In order to improve the data quality it is important to understand the barriers for it. After having identified the barriers for the data quality in the data migration they can be reflected to the methods to improve the data quality. The next chapter discusses and identifies these methods.

5.3 Improving master data quality in the data migration process

The interview themes seven and eight concentrated on improving the data migration process and especially improving the master data quality in the data migration process (later in this chapter referred to only as data quality). The ninth theme concluded these discussions as lessons learned type of summaries. This part of the interviews was the most important as it concentrated on the primary research question. Here the data quality barriers discussed in the previous chapter were reflected to the methods to improve the master data quality in the data migration.

According to the interviews what the business wants to achieve in its strategy is really the starting point in every project. In the most cases it is looking at not to have disparate systems and to start using an integrated system, such as SAP. Thus, the goal is to become more efficient and reduce costs of the operations.

The very first part is to illustrate to the customer the costs to the business of not having good quality data and raise the importance of data quality. It is essential to do this in the very beginning of the project because neglecting it in the first place may cause

difficulties later on. Then one usually gets a good buy in from the business and they understand that they actually need to put a little and in some cases lot of effort into that area. One interviewee also said that the data quality improvement process should be separated and emphasized as a part of the data migration process.

"What is critical is to get the buy in from the client that they understand the importance of the data." (I1.)

"When we talk about data migration project, we talk about the project which includes also data quality improvement project. We cannot just migrate the data because at the same time we have to improve the quality, which is difficult." (I5.)

Another thing is to emphasize the partnership approach. The client and the delivering team of consultants are working together towards the common goal. Therefore, it is not a traditional vendor-client relationship. In addition, one needs to have a good engagement with the client. That requires communication, understanding and trust. Also change management is tightly related to this as it is important that everyone within the organization understands that what will be done is better for the organization and better for them from the job security perspective because they end up with more efficient operative company. When everyone has bought the idea, it is half the battle. These things have a direct impact on the results of the project including the level of the data quality.

"I always emphasize that we are working as partners in delivering the end result which is moving to a new system with good quality data. It is definitely a partnership approach. They can't do it on their own; we can't do it on our own." (I1.)

In every project the most important part tends to be planning. However, according to the interviewees often planning is not on the required level. Thus, in order to succeed in the data migration one needs detailed plans for the schedule, roles, responsibilities, and project scope. Better planning could clarify the schedule which could lead to better data.

"The risks should be limited before migration. Therefore, one needs careful planning." (I8.)

Planning includes also the fact that the introduction and boarding for the project members should be arranged carefully. Also the client side needs to understand the project details and the data objects to be migrated. It is also problematic if people do not know the phases and where the project is going. Otherwise there will be lots of complicated questions which will delay the project and affect negatively also the data quality.

From a data perspective the plans should also include specific checkpoints or milestones for the data. There should be commonly agreed deadlines when the status of the data is revised. Therefore, the plans should be so detailed level that also the smaller data related checkpoints are agreed and communicated for all people at issue.

"The clear deadlines have direct impact on the process. If there are no deadlines, for example when the customer data needs to be cleansed, the process tends to slip from schedule." (I4.)

"The data quality can be ensured also during the bigger projects using the checkpoints." (I3.)

However, the checkpoints are unnecessary if one does not have clear rules or criteria which the data should fulfill. Therefore, the parameters for the data quality should be defined including how the data quality can be measured and proven in the project. There should be also detailed rules for the required content and format of the data, for example if capital letters are used in the customer names and in which format the phone numbers are stored.

"We need the rules for the content of the data and for the definitions of the fields so that they all look the same. Rules should be clear. No grey rules." (I7.)

It is also important that the system and the data in it reflect the business processes. Therefore, the users need to know what the business reasons behind the data fields are and what the values mean. Thus, it is always necessary for the data quality to clarify the business reasons. In addition, the data rules should be based on the business processes. The system should be also configured to correspond to the rules. For example, if some data field, such as customer address, is mandatory business-wise, it should be mandatory also technically in the system.

"Put the data in the system to reflect the business processes. First you need to define your business processes and then translate those to technical perspective." (I7.)

Apart from the data, rules can be assigned also to the processes. According to the interviews, there should be for example clear rules how to execute the data cleansing. Therefore, one should determine unambiguous steps and rules for the data cleansing process, how to identify obsolete data, and how to check duplicates. The rules can also be assigned to the data verification and validation. This has a major impact on the data quality.

An important part of the planning is that the roles and responsibilities are clear. The distribution of work needs to be clear and everyone should know their roles. It is important also for the data quality to define and communicate the data related roles and

responsibilities unambiguously. For example, it has to be clear who extracts, transforms, cleanses, uploads, and signs off each of the data objects. On the other hand, the data owners need to be known as well.

From data perspective it is very important who has the rights to create and maintain the data. The quality of the data is directly dependant on who controls it. In the system implementation projects there might be several persons who are loading and creating the data in the system. This can end up with having vast amounts of duplicates and poor quality data in the system. Thus, the general data should be maintained centrally in one place. Then it is possible to maintain the control of the data.

"It is important that the general a-level data is controlled centrally in a general way. Then the rules will always be the same." (I7.)

"For example, if all can create their own customer data, it is the basis for the problems" (I4.)

In every project there are obviously vast amount of files and documents. It was already discussed that the file handling and naming conventions can be a problem for the data migration and also data quality. Usually this is rather chaotic in several projects. Therefore, all the plans and processes should be documented. From the data perspective this means that all that is done to the data, such as cleansing and mapping, should be written down and documented. The documentation should be done in a common and agreed way so that everyone knows where the files and documents are saved and how there are named. Thus, the existence of rules applies also to the documentation and file handling.

"All the data spinnings and mappings have to be documented. And the documentation should be always done in the agreed place which is available for all the people at issue." (I6.)

Related to the file handling it is important that all the files that are used, such as data collection templates and data loading files, are clear and unambiguous. Files should also include the version history so that it is clear what is done in the file and when. All the used templates should be tailored and developed to fit to serve the project and its purpose. It is also important to carry out a clear walk through for the persons who use the templates.

"For example, the data collection templates could be developed so that they include all the necessary instructions, field descriptions, maximum lengths and the possible values." (I4.)

So the plans, rules, roles, and deadlines need to be transparent and communicated to all project members. Even more important is that the communication should be continuous between the project groups and members. For example, if the configuration changes, the functional team needs to communicate the changes for the data team. Otherwise, there will be problems with the data. On the other hand, it is important to have a clear goal from the beginning. Thus, the goal should be communicated for all project members from the beginning. From the data perspective this means that it should be clear also for the customer in which format the finalized data is in the target system.

As it was already discussed the major problems in the data migration are that data migration is seen only as a technical process and the data experts are included in the project only at the later steps. Therefore, in order the project to be successful all parties concerned should actually be there on the absolute start of the project. The project group should also be skillful enough. Lots of different knowledge is needed, such as technical, functional, and also business knowledge. The project needs experts for both legacy system and the target system. The business people should have also time to do the process related tasks. The best would be to release the project members from their day job to the project.

"The project needs skillful people who know the data they are working with. Also lots of functional skills are needed to have the data loaded in good quality." (I2.)

"The consultants need to be proven good. Technical skills, such as excel, sap tables, data dictionary, are needed. Otherwise it is quite difficult to do good data cleansing and to prove the customer that cleansing is good." (I8.)

"I can only state that the best projects have been when the people have been backfield from the day jobs to work full time on the project." (I1.)

In the previous chapter it was also mentioned that businesses do not want to spent money upfront for proper tools. According to one interviewee, people take a very short term view taking the applications. Nevertheless, tools are very important part of the data migration. There are tools which analyze the quality of data, control the flow of data and tools which will actually show the status of the progress within the project. Usually the tools are also rather simple, easy to use, and not too expensive. Therefore, the tools and applications available should be utilized. It is also possible to arrange a solution walkthrough in the beginning of the project for the customer of tools that can be appropriate for the project. Then the customer would know better which tools are useful and worth of paying.

According to the interviews an important method to improve the data quality is to analyze the status of the data in the legacy systems when the project starts. Before it can

be known how the data is migrated one needs to understand what he has and where to aim at. This means that the current data and the data structure need to be familiar before it can be transformed and restructured to fit to the target system. In addition, when the status of the data in the legacy system can be analyzed it is easier to prove the need for the improvement of the data quality.

"You need to be able to analyze their legacy data to be in the position to push back the condition of their current data. That actually helps to facilitate a greater buy in from the business. Once they can see some tangible reports you are really proving what you have been telling them from the start about the data quality." (I1.)

It was already discussed that the data cleansing and its execution is important for the data quality. One option to do it is to arrange the cleansing already in the legacy system. According to one of the interviewees this would be the best approach to cleanse the data. At the same time one should reduce the amount of the migrated data considerably by different criteria. No redundant data should be migrated. For example, the customer records which have not been used in years do not have to be migrated to the target system.

"Ideally you would clean everything in the source or as near the legacy system as possible. This leads to less iterations in the ETL process and easier reconciliation. Only necessary data needs to be migrated." (I2.)

In addition, the reconciliation and sign off methods should be clear. There is a difference between validating only ten percent and validating all data. In any case, it is important to validate the most important data properly. Also different tools can be used to reconciliation and validation, which should be used if available in the project.

"It is not easy to validate major amount of data. What is important is that you check the most important data, such as the most important customers, so that the addresses and others are correct. So you should control your main ones." (I7.)

Even though the thesis concentrates on the master data one interviewee stated that by improving the metadata quality one can improve also master data quality. Thus, metadata has a direct impact on the master data quality.

"Metadata means that the rules of the data are in place. It needs to be determined for example who owns the possible values of the data fields. The quality of the metadata affects the quality of the master data." (I6.)

An important phase in the data migration is testing. The testing concerns usually the system configuration and processes but it is as important to test the data. According to the interviews an efficient method to ensure the high quality data is by arranging

multiple test rounds for the data. The data should also be tested properly in the user acceptance testing and cutover simulation. Another method to improve data quality is to require sign offs for the data after every test round. Then the status of the data quality is monitored at regular intervals.

"Ideally there are at least three test uploads into different system environments for functional test, system integration test, and user acceptance test. The fourth upload is done into the production system. That is how the quality can be improved remarkably." (I3.)

In the interviews there were also different ideas how the data should be tested. Some said that the complete data sets should be tested and the other suggested quite the contrary. In any case, it was agreed that the testing is important for the data migration process.

"It is not important to load 90 percent of your data for tests. There is no use to load just for fun. It is sometimes better to have only a few customers for testing which represent different cases to test processes. So it is better to have representative customers little amount but good quality." (I7.)

"The data needs to be cleansed for the test rounds and as complete data sets should be tested as possible to catch all the technical issues." (I3.)

"It is a good practice to upload all the data also into the test environment." (19.)

One interviewee had also a different idea for the testing. It was a separate environment for the data related tests. There the data could be tested without affecting the others tests or processes.

"You might want to have a playground environment for data. There you could play with the data and would not affect all the others, such as solution or training. However, it is a luxury. But when talking about critical data you have to have the tools, such as separate environment, to test the data. This is one of the ideas which could improve the data quality." (I2.)

When talking about SAP, it is good to acknowledge the characteristics of SAP. Traditionally, SAP is very strict and requires a different approach for implementation than many other systems. According to one interviewee, if the customer respects standard SAP the implementation and migration work is easier and data quality is better. In other words, if the customizations and workarounds are minimized it is easier to implement the new system. Also SAP best practices were stated to be useful to follow.

"Try to convince customer to minimize the customization. If we respect the standard SAP and best practices, we will have better data quality." (I8.)

Finally, the table 5.2. concludes the interview results for this section. It lists all the methods to improve the master data quality in the data migration process.

Table 5.2. *The methods to improve the master data quality in the data migration process.*

The methods to improve the master data quality
Raise the importance of good quality data and get buy in from the business at the early stages of the project
Take care of change management
Utilize the tools and applications available
Release the project people from their day job to the project
Take care of the good engagement with the client
Emphasize the partnership with the client
Arrange a solution walkthrough for the customer of tools that can be appropriate for the project
At starting point analyze the status of the data in the legacy systems
Communicate the project scope, phases, schedule, roles etc. for all project members
Gather skillful project group (functional, technical and business knowledge, etc.)
Emphasize and facilitate the communication within the project team
Arrange and execute the data cleansing carefully already in the legacy system
Determine clear reconciliation and sign off methods for the data (use tools if available)
Include and engage all parties (incl. data people) in the project from the beginning
Determine rules for disciplined documentation (including file handling, naming conventions, change log, etc.)
In the user acceptance testing (UAT) test also data quality
Arrange a separate system environment (playground) for data related tests
Do multiple test uploads for the data to different systems
Require a sign off for the data after every test upload
Plan the data migration project, schedule, roles & responsibilities, etc. carefully and in detailed level
Arrange the introduction and boarding to the project team carefully
Create unambiguous data collection templates and carry out walkthrough for them
Reduce the amount of the migrated data considerably by different criteria (do not migrate redundant data)
Arrange scheduled checkpoints specifically for the data
Minimize the customization and workarounds in SAP (respect standard SAP solutions)
Comply with SAP best practices always when applicable
Define the parameters for the data quality (how the data quality can be measured and proven in the project)
Arrange the cutover simulation where data is also tested
Define and communicate the data related roles and responsibilities unambiguously (e.g. who

extracts, transforms, cleanses, uploads, and sign offs each of the data objects)
Clarify the goal in the beginning for the project members
Separate and emphasize the data quality improvement process as a part of the data migration process
Define unambiguous rules for the required content and format of the data (e.g. capital letters, phone number format, etc.)
Manage the general (a-level) data centrally in one place (creation, maintenance, etc.)
Always clarify the business reasons behind the data fields and values for the customer
Improve and clarify the quality of the metadata
Upload the whole data scope also to the test system and test complete data sets
Determine unambiguous steps and rules for the data cleansing process (e.g. what data is obsolete, how to check duplicates, etc.)
Determine SAP rules for data to correspond to the business rules (e.g. mandatory fields)
In test uploads and sign offs concentrate on validating only the most important data (e.g. most important customers)

In general, the interviewees agreed in many cases even though there were some contradictions with the perspectives for the testing. In addition, there were some ideas which were noted by only one interviewee. Therefore, it was useful to send the complete list of methods back to the interviewees in order to choose the most important methods. As a result from the additional questionnaire the table 5.3 lists these most intrinsic methods to improve the master data quality chosen by the interviewees. The interviewees were asked to select five of the most intrinsic methods but eventually after analyzing the answers six methods stood out from the original list.

Table 5.3. *The most intrinsic methods to improve the master data quality in the data migration process.*

The most intrinsic methods to improve the master data quality
Take care of the good engagement with the client
At starting point analyze the status of the data in the legacy systems
Arrange and execute the data cleansing carefully already in the legacy system
Create unambiguous data collection templates and carry out walkthrough for them
Define and communicate the data related roles and responsibilities unambiguously (e.g. who extracts, transforms, cleanses, uploads, and sign offs each of the data objects)
Determine SAP rules for data to correspond to the business rules (e.g. mandatory fields)

In all, the methods to improve the master data quality varied highly depending on the viewpoint for the data migration. Some methods were more high-level courses of action and suitable almost for any project, e.g. careful planning. However, some of the methods gave more detailed insights into the data migration. These will be discussed more in the chapter six.

5.4 The prospects

The final interview theme concentrated on the prospects of the studied phenomenon in general. The discussions varied from the future of the ERP systems, and especially SAP systems, to the future technologies and how they affect the data.

Many interviewees acknowledged that the amount of the traditional new ERP system implementations will probably fall in the coming years. Instead there will be lots of business transformation projects when the businesses update and enhance the systems and processes. The interviewees also agreed that the future of SAP looks bright even though new competitive technologies require new innovations.

"In the future the principals remain the same. Not so much new SAP implementations anymore but transformation projects and merging projects." (I7.)

From the new technologies especially cloud, mobility and data analytics were said to be most evolving ones. The cloud storages are already widely used and in the future they will be even more popular as it saves money when businesses don't have to invest and work with the hardware. Also mobility will grow in the future when even more software providers are starting to provide also mobile applications for phones and tablets.

According to the interviewees the importance of the data is becoming increasingly important. Data is essential for the companies to run their business. It is collected from everywhere in variety of formats and it is created all the time even more. Therefore, powerful data analytic techniques and tools are also required. There also lays a question how all the data can be exploited. However, all the interviewees agreed that the meaning of the data will grow.

"Data is the new oil." (I5.)

New technologies may also make it easier to handle data. Nevertheless, it does not remove the fact that data still needs to be handled in IT systems and implementation projects. One interviewee stated that data handling may become even more complex when the volume of the data grows, velocity gets faster, and variety expands in the future.

All in all, the future is sort of a mystery. Technology enables new things so fast that it is almost impossible to see the coming possibilities. What is certain is that the data is not going anywhere. Thus, it is important to accept the significance of the data for the businesses and really understand the meaning of good data quality.

6. DISCUSSION

The interviews and the additional questionnaire provided lots of material and insights into the studied phenomenon. This chapter concentrates on discussing the most important results in more detailed level. The results are observed from different perspectives and reflected in contrast to the theory. The structure of this chapter follows the structure of the previous chapter where the results were presented. The interviews focused especially on the SD master data but in this chapter it is often referred to only as data.

6.1 Data quality and perfect data quality

The discussion about data and data quality formed the foundation for the whole study. Generally, it is obvious that data is important for business. As it was stated already in the chapter two data is the key business asset hence it needs to be actively managed to ensure it is accurate, current and fits for purpose (Marsh 2005, p. 106). It is also a fact that data is the foundation for operational, tactical and strategic decisions of companies (Madnick et al. 2004).

According to the interviews the importance of data for businesses was well known among the interviewees. As one of the interviewees said: "Data is an asset, data is a business enabler". Or as another stated: "…no business process runs without data." These citations describe well how the purpose of data is seen in the business world.

What comes to the definition of data quality the multidimensional concept of data was also well understood. Even though it was understood the different dimensions were not so familiar. The dimensions which were often mentioned in the interviews were correctness, accuracy, completeness, reliability, cleansed, and timeliness. However, as it was described in the chapter two there are dozens of data quality dimension which can be used to manage and measure the data quality in organizations (McGilvray 2008, p. 7).

The theory review revealed that a classic definition of data quality is "fitness for use (Tayi & Ballou 1998, p. 54; Haug & Arlbjørn 2011, p. 292). Thus, the data quality depends on the purpose of the data. This was also commonly understood in the interviews. It was stated: "Data quality is good when the data serves its purpose…".

When talking about the required level of the data quality, it came clear that no organization needs, wants or will pay for perfect quality data (Morris 2012.) This was agreed also in the interviews as “perfect data has also a prefect price” which organizations are not willing to pay. Instead the resources should be focused especially on the most important data objects from the business perspective. The data objects depend on the business but generally the most important ones are customers, vendors and materials. From the sales and distribution perspective the most important are customers and materials.

The general data quality problems were also discussed with the interviewees. To conclude the general data quality issues from the chapter two, too much data is just plain wrong, too hard to find, poorly defined, inconsistent with other data, and at risk of being lost or stolen (Redman 2008). According to the interviewees, the problems are usually quite similar from organization to organization and from project to project. Data-wise there are always duplicates, invalid data, and incorrect data to cleanse. The data quality issues in the context of data migration are discussed in more detailed level in the next chapter.

It can be concluded that data and data quality in general were well known among the interviewed consultants. The characteristics, purpose, and importance of data were understood which formed a great foundation for the discussion about data quality in data migration.

6.2 Data quality barriers in the data migration process

The interviews provided a long list of barriers in the data migration. In order to get more out of the results the data quality barriers were divided into three different groups: data, people, and process related barriers. However, as the good quality data is the key element and the main deliverable of any data migration process all the barriers in the data migration process can be seen as challenges for the data quality.

As the data quality barriers were discussed especially from the master data perspective, it is obvious that all the barriers affect the master data quality. Nevertheless, the barriers can be applicable for the other types of data as well.

This chapter discusses the data, people, and process related data quality barriers in the data migration process in the separate subchapters.

6.2.1 Data related barriers

The data related barriers are directly related to the data itself and the rules, systems, and techniques behind the data. The data related data quality barriers from the interviews are described in the table 6.1.

Table 6.1. *The data related data quality barriers in the data migration process according to the interviews.*

The data related data quality barriers
The importance of data is underestimated
Data is not seen as an investment
Master data is seen only as a static data
Huge volumes of data
The legacy system infrastructure is not documented at all
The data in the legacy system is poor quality
Businesses do not know the status and quality of the data in the legacy system
Businesses don't understand their data
Data models in the legacy system and the target system are not consistent
Data conversion is hard from the legacy to the target
The business reasons for the data requirements are unclear
No clear data quality rules or criteria
Data uploading techniques are challenging
Poor technical tools
The data is not tested
Everyone has access and possibility to handle the data

When comparing the barriers raised by the different interviewees with the different background and experience, there really was not any correlation between the background, experience, and some certain type of barriers. In all, the discussions with the different interviewees were quite different and most of the interviewees had some factors which only they raised as a data quality barrier in the data migration.

However, there clearly were some barriers which were mentioned by several interviewees. In fact, the following data related data quality barriers were mentioned by more than two interviewees: the data in the legacy system is poor quality, data conversion is hard from the legacy to the target, businesses do not know the status and quality of the data in the legacy system, huge volumes of data, and no clear data quality rules or criteria.

The barriers in the data migration process raised from the theory were discussed already in the chapter 3.2.3. In general, many of the barriers from the theory corresponded to the data quality barriers raised in the interviews. For example, the data related barriers came up more or less in both theory and interviews. Some of these barriers were the importance of the data quality not being understood, lack of confidence in the data, lack of master data control routines, large data volume, poor master data quality, businesses who don't understand their data and don't know how to fix the data quality, different data models in the source and the target system, error prone master data capture and processing, data quality control, and proper testing, (e.g. Morris 2012; SAP 2013b; Haug & Arlbjørn 2011; Matthes et al. 2011; Haller et al. 2012; Knolmayer & Röthlin 2006; Xu et al. 2002; Redman 2008).

However, the barriers which were found from the theory but did not come up in the interviews are a target database already containing data, poor failure detecting, insufficient reconciliation, challenging to ensure the quality of the metadata, data loss, data completeness, data semantics, data corruption, the fact that measuring data quality is problematic, poor data definition, and data privacy and security (e.g. Matthes et al. 2011; Haller 2009; Redman 2013; Morris 2012; Redman 2008).

On the contrast, some barriers which raised in the interviews did not come up in the theory review. These barriers are the fact that the master data is seen only as a static data, undocumented legacy system infrastructure, unclear business reasons for the data requirements, challenging data uploading techniques, poor technical tools, and the problem if everyone has access and possibility to handle the data.

The data related barriers give good understanding about the general barriers for the data quality in the data migration. However, not all the barriers are directly related to data. The next chapter focuses on the people related data quality barriers in the data migration.

6.2.2 People related barriers

The people related barriers are indirectly related to the data but directly related to the people in the data migration process. These barriers include communication, people's roles, human behavior, and people's knowledge and experience related challenges. The people related data quality barriers raised in the interviews are listed in the table 6.2.

Table 6.2. *The people related data quality barriers in the data migration process according to the interviews.*

The people related data quality barriers
Weak buy in for data quality from the business
The role of the data team is underestimated
Data experts are included in the project only at the later steps (not from the beginning)
The fear of the change
Inefficient management of change
Project people do not have commitment for the project
There is no communication within the project team or it is hard, e.g. difficult terms, jargon, etc.
The data is siloed for different people and units
Difficult customers
Poorly trained or unskillful project team
The systems are not familiar enough for the project team
Human errors
Roles, ownerships and responsibilities are not clear within the project team
Not operating master data governance team for the legacy system

Regarding the people related data quality barriers there wasn't any correlation between the background, experience, and some certain type of barriers. However, the barriers which stood out from the list being mentioned more than once in the interviews were the following: there is no communication within the project team or it is hard, e.g. difficult terms, jargon, etc., roles, ownerships and responsibilities are not clear within the project team, inefficient management of change, not operating master data governance team for the legacy system, and poorly trained or unskillful project team.

The people related barriers which came up in both theory and interviews are lack of change management, lack of commitment to the project, poor communication, inadequate training, lack of experience, unskillful project team, the systems not being familiar to the project team and they not being user friendly, lack of roles and responsibilities, and lack of maintenance of the master data (e.g. Hawking et al. 2004; Xu et al. 2002; Matthes et al. 2011; SAP 2013b; Umar et al. 1999; Haug and Arlbjørn 2011).

On the other hand, there were again some barriers which raised from the theory but not in the interviews. These barriers are the problem with identifying the right partner, people creating the data having little understanding how the data is used in the organization, challenging to get the creators of data to partner with the users of the data, lack of discipline, misplaced benefit ownership, lack of rewards, lack of data quality owners, organization structure, employee relations, the time external consultants

required to be in the project, and the fact that people cannot find the data they need (e.g. SAP 2013b; Redman 2013; Hawking et al. 2004; Umar et al. 1999; Xu et al. 2002; Redman 2008).

Finally, the people related barriers which came up in the interviews but not in the theory are weak buy in for data quality from the business, the underestimated role of the data team, the problem that the data experts are included in the project only at the later steps, the fear of change, the data being siloed for different people and units, difficult customers, and human errors.

People have important roles in any data migration project. However, human behavior and the characteristics of the people highly affect the success of the project. Nevertheless, in addition to the data and people related barriers there are other barriers for the data quality in the data migration. These barriers are discussed in the next chapter.

6.2.3 Process related barriers

The process related data quality barriers are indirectly related to the data but directly related to the processes in the data migration, such as planning and resources. This group combines all the barriers which are not directly related to the data and the people in the data migration. The process related data quality barriers from the interviews are listed in the table 6.3.

Table 6.3. *The process related data quality barriers in the data migration process according to the interviews.*

The process related data quality barriers
Data migration is seen only as a technical process
Data migration process is not understood
The requirement for improving data quality is a surprise
The data migration project is planned poorly
Data migration process and the schedule is underestimated
The data migration project scope is not clear
The project does not have enough resources
Surprising changes in the project, system etc.
Businesses do not want to spent money upfront for proper data tools etc.
Documentation and file handling is not disciplined
Incorrect or unclear rules and definitions for business processes
It is hard to translate the business requirements to technical terms

The following process related data quality barriers stood out from the full list: data migration process and the schedule is underestimated, the data migration project is planned poorly, data migration is seen only as a technical process, the data migration project scope is not clear, and documentation and file handling is not disciplined. These barriers were mentioned by more than one interviewee. Again there was no correlation between the background, experience, and some certain type of barriers.

Again some of the barriers raised both in the interviews and from the theory. These barriers are the fact that the data migration is often seen only as a technical issue, unrealistic and poor planning, delays in the project, poor resources, and a need to adapt the business processes (e.g. Morris 2012; SAP 2013b; Haller et al. 2012; Eckerson 2002, Matthes et al 2011; Hawking et al. 2004; Mandal & Gunasekaran 2003).

The barriers which came up only in the theory but not in the interviews are the challenge of coordinating all data migration programs and invoking them in the correct order, project profitability, reputation, and regulation, target application stability, orchestration, dimensioning, interference, inadequate process engineering and ongoing support, poor business performance, poor application management, neglecting administrative details, and organizational confusion (e.g. Matthes et al. 2011; Hawking et al. 2004; Umar et al. 1999; Redman 2008).

On the other hand, the process related barriers which came up only in the interviews are the fact that the data migration process is not often understood, the requirement for improving data quality is a surprise, unclear project scope, surprising changes in the project or the system, businesses not wanting to spent money upfront for proper data tools, documentation and file handling not being disciplined, and the fact that it is hard to translate the business requirements to technical terms.

To conclude, the interviews concentrated especially on the SD master data. However, all the barriers mentioned can also be seen as barriers for the general data quality in the data migration. Nevertheless, the interview results and the theory regarding the barriers in the data migration corresponded only partly to each other. In all groups, data, people, and process related barriers, there were some barriers which came up only in the theory. On the other hand, in all groups there were also some new barriers which came up only in the interviews.

The main reason for the differences is that the theory approached the barriers for the data migration from different perspectives, such as general barriers for the data migration, ERP implementation barriers, and data quality barriers in general, whereas the focus in the interviews was especially data quality barriers in the data migration for the master data. The consultants' perspective in the interviews gave also some fresh insights according to their experiences from the data migration.

6.3 Improving master data quality in the data migration process

The methods to improve the master data quality in the data migration process were derived from the data quality barriers in the data migration process discussed earlier. These methods were also divided into the same three groups which were used in the previous chapter: data, people, and process related methods.

This chapter discusses the data, people, and process related methods to improve the master data quality in the data migration process in the separate subchapters (later in this chapter referred to often as data quality). The final subchapter also points out the most intrinsic methods which distinguished in the additional questionnaire.

6.3.1 Data related methods

The data related methods are directly related to the data itself and the rules, systems, and techniques behind the data. The data related methods to improve the master data quality in the data migration are described in the table 6.4.

Table 6.4. *The data related methods to improve the master data quality in the data migration process.*

The data related methods
At starting point analyze the status of the data in the legacy systems
Arrange and execute the data cleansing carefully already in the legacy system
Reduce the amount of the migrated data considerably by different criteria (do not migrate redundant data)
Define unambiguous rules for the required content and format of the data (e.g. capital letters, phone number format, etc.)
Define the parameters for the data quality (how the data quality can be measured and proven in the project)
Always clarify the business reasons behind the data fields and values for the customer
Determine SAP rules for data to correspond to the business rules (e.g. mandatory fields)
Determine unambiguous steps and rules for the data cleansing process (e.g. what data is obsolete, how to check duplicates, etc.)
Create unambiguous data collection templates and carry out walkthrough for them
Arrange a separate system environment (playground) for data related tests
Do multiple test uploads for the data to different systems
Upload the whole data scope also to the test system and test complete data sets
In the user acceptance testing (UAT) test also data quality
Arrange the cutover simulation where data is also tested
Require a sign off for the data after every test upload

Determine clear reconciliation and sign off methods for the data (use tools if available)
In test uploads and sign offs concentrate on validating only the most important data (e.g. most important customers)
Arrange scheduled checkpoints specifically for the data
Manage the general (a-level) data centrally in one place (creation, maintenance, etc.)
Improve and clarify the quality of the metadata

When reflecting the methods to improve the master data quality to the barriers discussed earlier, these methods can overcome some of the barriers. For example, the barriers which have been taken into consideration in the methods are poor quality data in the legacy system, businesses not knowing the status and quality of the data in the legacy system, inconsistent data models in the legacy system and the target system, data conversion being hard from the legacy to the target, unclear business reasons for the data requirements, no clear data quality rules or criteria, untested data, and everyone having access and possibility to handle the data.

On the other hand, the barriers which cannot be overcome by the methods listed above are underestimated importance of data, data not seen as an investment, master data seen only as a static data, huge volumes of data, the legacy system infrastructure not being documented, businesses not understanding their data, challenging data uploading techniques and poor technical tools.

The methods took also into consideration some barriers which did not raise in the interviews but only in the theory. These barriers are for example insufficient reconciliation, challenges in ensure the quality of the metadata, and the fact that measuring data quality is problematic (e.g. Haller 2009; Redman 2013; Morris 2012).

When considering the data migration process in SAP implementation, it includes six steps: analyze, extract, clean, validate, load, and reconcile (SAP 2014a). Most of the data related methods to improve the master data quality are directly related to the migration process steps. For example, one method states that at starting point the status of the data should be analyzed in the legacy system. According to the analysis the quality of the source data can be assessed and the needs to cleanse and normalize the data can be clarified (SAP 2014a).

The method to create unambiguous data collection templates and carry out walkthrough for them is related to the extraction phase. Usually the data is extracted to a so called staging area where it can be prepared for the target system (SAP 2014a). The staging area can be a specific tool or just a data collection template where the data is collected and processed for further use.

Some of the methods are directly related to the cleansing phase. These methods are determining unambiguous steps and rules for the data cleansing process, defining unambiguous rules for the required content and format of the data, reducing the amount of the migrated data considerably by different criteria, and arranging and executing the data cleansing carefully already in the legacy system. The cleansing can be done in the staging area after the extraction or as Haller et al. (2012, p. 172) state it can also be beneficial to execute the data cleansing already in the source system. It can simplify the migration in the latter steps as it reduces the data processing after the extraction and helps the reconciliation between the target and the source systems.

The fourth phase in the data migration is the data validation against the business rules and the target system context. This also includes the transformation process to transform the data to meet the requirements and the data model of the target system. (SAP 2013b.) In addition, a data collection template can be helpful in transforming the data towards the target system data model as the collection template reflects the model in the target system. The method which is also related to this phase is clarifying the business reasons behind the data fields and values for the customer. When validating and transforming the data, the business rules behind the data need to be known. At the end of the day, the data and the system should reflect the business rules and support the daily business without problems or interruptions. Therefore, also the method to determine SAP rules for the data to correspond to the business rules is important.

An important part of loading is test loads. Therefore the methods related to the loading phase are to do multiple test uploads for the data to different systems, upload the whole data scope also to the test system and test complete data sets, require a sign off for the data after every test upload, and in test uploads and sign offs concentrate on validating only the most important data. It should be noted that all the methods cannot be exploited in parallel as the methods include also contradictions. For example, one method states that test uploads and sign offs should be concentrated on validating only the most important data. On the other hand, the other method state that the whole data scope should be uploaded into the test system and the complete data sets should be tested.

The final step in the data migration is reconciliation. The method related to that is determining clear reconciliation and sign off methods for the data. The reconciliation is important to verify the results and successful migration (Haller 2009). It is important to determine who is the correct person to reconcile the results and what is the best method to execute it.

Apart from the data migration process phases mentioned earlier measuring the data quality is important as well in order to justify the improved quality of the data. Therefore, it is essential to define the parameters for the data quality and determine how the data quality is measured and proven in the project. If you cannot measure it, you

cannot manage it (Bobrowski 1999, p.4). Thus, it is highly useful to recognize the various data quality dimensions as they can be used to measure and manage data quality (McGilvray 2008, p. 7). In addition, it is highly useful to arrange scheduled checkpoints specifically for the data, where the quality of the data and the progress of the migration can be measured.

Testing is an important part of the general ERP implementation and data migration process. Different types of tests are for example unit testing, integration testing and user acceptance testing (Morris 2012). From data quality perspective it is important to test also data quality. Therefore, testing related methods to improve the data quality are to test also data quality in the user acceptance testing and arrange a separate cutover simulation where data is also in an important role. Then the actual users have possibility to use the actual data and feed back any data related problems. One approach to improve the data quality is to arrange a separate system environment only for data related tests where the data could be tested without affecting any other tests.

From the master data management perspective it is important to retain the control over the creation and maintenance of the data. It is possible only if the management is done in a controlled way. Therefore, in the data migration process the general data should be managed and created centrally in order to retain the control.

McGilvray (2008) stated that the quality of the metadata impacts all the other data categories. As it was described in the chapter two for example documenting definitions in the metadata improves quality because it transforms undocumented assumptions into documented and commonly agreed meanings (McGilvray 2008, p. 44.) Therefore, in order to improve the quality of the master data it is also important to improve the quality of the metadata.

When considering the methods which stood out in the interviews the following distinguished from the list: create unambiguous data collection templates and carry out walkthrough for them, determine unambiguous steps and rules for the data cleansing process, and arrange scheduled checkpoints specifically for the data. These methods were mentioned by three interviewees. On the other hand, when considering the correlation between the background, experience, and some certain type of methods, there was no clear pattern.

In all, the data related methods to improve the master data quality are highly related to the data migration process phases. They also took into account the important fact that in order to improve the quality one needs to be able to measure the quality. It was also mentioned that the data should reflect the business rules and processes, which is in line with the fact that the quality of the data is good when it serves its purpose. However, the

data quality can also be improved by methods not directly related to data. The next chapter discusses about the people related methods to improve the data quality.

6.3.2 People related methods

The people related methods are indirectly related to the data quality but directly related to the people in the data migration process. These methods include communication, people's roles, human behavior, and people's knowledge and experience related procedures. The people related methods to improve the master data quality in the data migration are listed in the table 6.5.

Table 6.5. *The people related methods to improve the master data quality in the data migration process.*

The people related methods
Raise the importance of good quality data and get buy in from the business at the early stages of the project
Take care of change management
Take care of the good engagement with the client
Emphasize the partnership with the client
Release the project people from their day job to the project
Gather skillful project group (functional, technical and business knowledge, etc.)
Emphasize and facilitate the communication within the project team
Arrange the introduction and boarding to the project team carefully
Include and engage all parties (incl. data people) in the project from the beginning
Define and communicate the data related roles and responsibilities unambiguously (e.g. who extracts, transforms, cleanses, uploads, and sign offs each of the data objects)

When considering the methods and the barriers together, the barriers which have been taken into consideration in the methods are weak buy in for data quality from the business, underestimated role of the data team, data experts are included in the project only at the later steps, the fear of the change, inefficient management of change, lack of communication within the project team or it is hard, poorly trained or unskillful project team, unfamiliar systems for the project team, difficult customers, and unclear roles, ownerships and responsibilities within the project team

On the contrast, the barriers which have not been taken into consideration are project people who do not have commitment for the project, the data is siloed for different people and units, human errors, and no operating master data governance team for the legacy system.

Apart from the barriers raised in the interviews the barriers in the ERP implementation projects are highly related to the people related methods to improve the master data quality. According to Hawking et al. (2004) ERP implementations have several barriers, such as lack of change management, poor communication, inadequate training, inadequate internal staff, misplace benefit ownership. The people related methods, such as take care of change management, gather skillful project group, emphasize and facilitate the communication within the project team, and define and communicate the data related roles and responsibilities unambiguously, take into consideration the ERP implementation barriers.

On the other hand, also some data related barriers, such as some managers refuse to admit that their data is not good enough and the importance of data quality is not understood (e.g. Morris 2012, Redman 2013), are also taken into consideration in the people related methods. For example, the method to raise the importance of good quality data and get buy in from the business at the early stages of the project is applicable to overcome these barriers.

The more general methods, such as taking care of the good engagement with the client, emphasizing the partnership with the client, arranging the introduction and boarding to the project team carefully, and including and engaging all parties in the project from the beginning, enhances the culture and the spirit within the project team. It also improves the communication in the project and makes it easier to manage. In addition, if the project people are released from their day job to the project, they are eager to participate and execute the project related tasks more carefully.

The most mentioned people related methods to improve the master data quality in the data migration are defining and communicating the data related roles and responsibilities unambiguously, gathering skillful project group, and including and engaging all parties (incl. data people) in the project from the beginning. Three interviewees emphasized these methods. Again there was not any clear correlation between the background, experience, and some certain type of methods.

People related methods to improve the master data quality in the data migration are more general in nature and can be applicable to any project. This is understandable because people are the key resource in any project. Even though they don’t have direct impact on the data, they affect indirectly the data quality. This applies also to the process related methods which are discussed in the next chapter.

6.3.3 Process related methods

The process related methods are indirectly related to the data but directly related to the processes in the data migration, such as planning and resources. This group combines

all the methods which are not directly related to the data and the people in the data migration. The process related methods to improve the master data quality in the data migration are listed in the table 6.6.

Table 6.6. *The process related methods to improve the master data quality in the data migration process.*

The process related methods
Plan the data migration project, schedule, roles & responsibilities, etc. carefully and in detailed level
Clarify the goal in the beginning for the project members
Separate and emphasize the data quality improvement process as a part of the data migration process
Communicate the project scope, phases, schedule, roles etc. for all project members
Determine rules for disciplined documentation (including file handling, naming conventions, change log, etc.)
Minimize the customization and workarounds in SAP (respect standard SAP solutions)
Comply with SAP best practices always when applicable
Arrange a solution walkthrough for the customer of tools that can be appropriate for the project
Utilize the tools and applications available

When reflecting the methods to the barriers, there are some common factors. The barriers which have been taken into consideration are the data migration process not being understood, the requirement for improving data quality coming as a surprise, poor planning, underestimated data migration process and schedule, unclear data migration project scope, businesses not wanting to spend money upfront for proper data tools, and documentation and file handling not being disciplined.

On the other hand, the barriers which have not been taken into consideration are data migration seen only as a technical process, the project not having enough resources, surprising changes in the project, system etc., incorrect or unclear rules and definitions for business processes, and it difficulties in translating the business requirements to technical terms.

Most of the data migration barriers, such as lack of roles and responsibilities and lack of scheduling scenarios (e.g. Haug and Arlbjørn 2011; Umar et al. 1999), can be overcome by more careful planning. This is rather general method but important is to plan the migration process in detailed level including the schedule, roles and responsibilities. In addition, the plan including the project scope, phases, schedule, roles, etc. should be communicated for all project members. As a part of this the project goal should be communicated and clarified for the project members in the beginning.

When discussing about SAP implementation, SAP is a complex system and it needs a different approach for its implementation than many other systems (Lau 2005). To ease the whole process it would be beneficial to comply with SAP best practices and respect the standard SAP solutions. When minimizing the customization and workarounds, the data migration would also be easier and it would affect the data quality as well.

Some process related methods take into consideration some data related barriers as well. For example, the method to separate and emphasize the data quality improvement process as a part of the data migration process would raise the understanding about the data migration process and the data quality in general. This would take into consideration the issue that importance of data quality is not understood (e.g. Morris 2012).

Technical perspective is also important to the process related methods. According to the interviews the tools to ease many steps in the data migration are available but they are not known or businesses don't want to spend money upfront for proper data tools. Therefore, it would be beneficial to arrange a solution walkthrough for the customer of tools that can be appropriate for the project.

Apart from the tools the projects usually include a vast amount of different files and version of the same files. Therefore, the rules for disciplined documentation should be determined and communicated so that for example file handling, naming conventions, and change logs would be maintained correctly for all files.

Also regarding the process related methods to improve the master data quality there wasn't any correlation between the background, experience, and some certain type of methods. However, the methods which stood out from the list were the following: plan the data migration project, schedule, roles & responsibilities, etc. carefully and in detailed level and determine rules for disciplined documentation. The methods were mentioned by three interviewees.

As the people related methods, also the process related methods are more general in nature. However, they are still important from the data quality perspective. Finally, the next chapter discusses about the most intrinsic methods identified from all the methods discussed earlier.

6.3.4 The most intrinsic methods

The additional questionnaire helped to identify the most relevant methods to improve the master data quality in the data migration. The table 6.7. describes these methods divided into the groups described earlier.

Table 6.7. *The most intrinsic methods to improve the master data quality in the data migration process divided into the focus groups.*

The most intrinsic methods to improve the master data quality	**Group**
Take care of the good engagement with the client	People
Define and communicate the data related roles and responsibilities unambiguously (e.g. who extracts, transforms, cleanses, uploads, and sign offs each of the data objects)	People
At starting point analyze the status of the data in the legacy systems	Data
Arrange and execute the data cleansing carefully already in the legacy system	Data
Create unambiguous data collection templates and carry out walkthrough for them	Data
Determine SAP rules for data to correspond to the business rules (e.g. mandatory fields)	Data

As it can be seen in the table 6.7. two of the most intrinsic methods are people related methods and the rest are directly data related methods. The process related methods did not stand out in the questionnaire. The recognized most intrinsic methods can be utilized in parallel in the same project as they are not in contradiction with each other.

Good engagement with the client is related to the people and interaction between them. It is important because it is the foundation for the good relationship. It is based on trust and efficient communication. When the engagement is good it is easier to move on with the project and communicate and solve also the difficult issues. This also applies to the data quality improvement and challenges related to that.

Roles and responsibilities are important in every project. In the data migration the importance stands out as the data migration includes several different phases which affect the data and the data quality. It has to be clearly decided and understood who will extract the data from the legacy system, cleanse it, transform it to correspond to the target system data model, upload it to the target system, and validate and sign off the data objects. When the responsibilities are clear from the beginning the people can prepare themselves for it and be involved in the process right from the start. This will have a positive impact on the progress and the results of the data migration process.

Often the data quality in the legacy system is poor. Therefore, it is important to analyze it in order to understand the actual status of the data quality. When creating tangible reports of the data quality status, it is also easier to convince the client and the project team that the data quality needs improvement. Analyzing the data in the legacy system also helps to understand the data model in the legacy system which in turn helps the transformation of the data from the legacy data model to the target data model.

The data cleansing can be done in different steps and in different ways. It is especially important to remove the duplicates and invalid data, correct the errors in the data and fill in the missing data. When this is done already in the legacy system it makes the rest of the process easier. For example, in the transformation phase one does not have to concentrate on the data cleansing anymore if it is done already in the legacy system. This makes it easier to reconcile the final data in the target system against the original data in the legacy system.

The data collection can be carried out in different ways depending on the project and the data object. But especially the master data can be collected using the data collection templates which are planned especially for the specific project, data object, and purpose. The templates should take into consideration things such as the mandatory fields, field descriptions, field lengths, possible values, and descriptions for the possible values. In addition, the template should correspond to the business rules behind the data and SAP structure for the data. These collection templates should be self-explaining and unambiguous so that the issues in the data collection could be overcome.

The system and the data should support the business processes. Therefore, also the technical rules in the system should reflect the business rules. From the data perspective especially the data related SAP rules should correspond to the business rules. For example, the mandatory data fields from the business perspective should be also mandatory in the system. This has a direct effect on the data quality.

When comparing the identified most intrinsic methods to the methods which were mentioned in the several interviews, only two of them were mentioned by more than two interviewees. These were defining and communicating the data related roles and responsibilities unambiguously and creating unambiguous data collection templates and carrying out walkthrough for them. In fact three of the methods were mentioned by only one interviewee: arrange and execute the data cleansing carefully already in the legacy system, determine SAP rules for data to correspond to the business rules, and take care of the good engagement with the client. Thus, it can be concluded that it was very beneficial to arrange the additional questionnaire in order to identify the most important results.

All in all, the methods raised in the interviews included most of the barriers identified in the interviews. Even though the interviews concentrated on SD master data some of the methods, especially people and process related, are very high-level advices which in any case should be taken into consideration in the data migration process in order to improve the master data quality. The methods don’t either give detailed instructions on how they can be utilized. Nevertheless, the methods clearly indicate how the master data quality can be improved in the data migration.

6.4 The prospects

SAP has long traditions providing software systems for the business use. ERP systems have for years been one of the most implemented IT solutions. On the other hand, traditional ERP systems are confronting more and more competition from the developing techniques such as cloud based solutions. Thus, it was interesting to discuss how the interviewees see the future of data migration and data quality.

In all, the discussion remained very high level and the interviewees did not have much to say about the subject. The main themes were the evolving techniques which will affect the very foundations of the ERP systems and data processing. Also the fact that almost all new ERP implementation projects are done came up in many discussions. In the future the ERP projects will be merely business transformation projects when the businesses update and enhance the systems and processes. In addition, there are always system consolidation programs generated by business merging. (Morris 2012, p. 1.)

However, the most important message from the final theme was that the data is not going anywhere. If something, the meaning of the data will grow. Therefore, it is essential to understand the concept of the data and learn how to manage it.

7. CONCLUSION

This chapter compresses and concludes the research. Firstly, the results of the research are summarized in relation to the research questions. Secondly, the chapter includes the assessment of the conducted study and discusses the potential further research.

7.1 Summary of the results

The research concentrated on the data migration in an ERP implementation project. The main interest in that phenomenon was master data quality. The primary research question was defined as:

- How to improve the master data quality in the data migration of SAP ERP system implementation?

The supporting secondary questions were:

- What is meant by good data quality?
- What is master data?
- What is data migration as a part of ERP implementation?
- What are the barriers for the data quality in the data migration process?

The research was conducted at a general level and specific or unique cases or organizations were not examined. The purpose was to compile a general list of methods how to improve the master data quality in the studied phenomenon. The study was conducted keeping in mind the organizations who are about to renew their ERP system and are concerned about the master data quality. In addition, the empirical part of the research studied only the consultants' perspective. The customers' point of view was not studied at all.

Next, the results for the secondary research questions are summarized in separate sections. Finally, the results for the primary research question are covered.

What is meant by good data quality?

Data constitutes the raw material for the Information Age. It is an asset and a business enabler. It is the key element for all information systems. No business process runs without data. Unlike physical raw material data is not consumed and can be reused

repeatedly for various purposes. In addition, unlike other typical resources data can be copied, shared and combined in many ways, it multiplies, it creates value when it is on the move, it is not consumed with use, and it is intangible.

A classic definition of data quality is "fitness for use". Therefore, the data quality depends on how the data is used. The data quality can be seen as a contextual concept. Thus, the data can be meaningless unless placed in some context. In addition, the data with the quality considered appropriate for one use may not possess sufficient quality for another use. However, no organization needs, wants or will pay for perfect quality data. Perfect data has also a perfect price. On the contrast, the data quality should be sufficient in regard to its purpose.

On the other hand, the data quality is a multidimensional concept. Thus, the data quality can be divided into several dimensions, such as accuracy, timeliness, completeness, correctness, meaningfulness, reliability, security, etc. It is useful to recognize the different dimensions as they can be used to manage and measure the data quality.

Organizations have often troubles with the data and data quality. Too much data is just plain wrong, too hard to find, poorly defined, inconsistent with other data, and at risk of being lost or stolen. In addition, organizations do not know the data they have, redundancy is out of control, and too much data are never used for anything.

To conclude, the data quality is good when the data serves its purpose and doesn't cause any interruptions for the business. It can be measured and managed by the means of the different data quality dimensions.

What is master data?

Data can be divided into different categories. One of these categories is master data. It is the core data and the main component of most information systems. Characteristic for master data is that it is created once, used many times and does not change frequently. In ERP systems master data is usually used across several application modules in the transactions that are registered.

Master data describes the people, places, and things that are involved in an organization's business. For example, customer data, employee data, vendor data and, product data are traditional master data in the organizations.

This study concentrated especially on the sales and distribution (SD) related master data. The SD master data refers to a centralized data object that stores information related to the sale of products and services of an organization in the SD module of SAP ERP system. It reduces data redundancy and makes more efficient the transactions

related to the customers, business partners, and the goods and services of an organization. The SD master data is basically the data about the customers, materials, and the conditions on which sales depend.

All in all, the master data quality is absolutely required for an ERP system to function properly. Because the transactions use the master data, if the master data is not correct, the transactions do not fulfill their intended purpose. Even a small amount of incorrect master data can result in great costs. In addition, the master data might be the one piece to SAP ERP implementation that can make the difference between getting the expected results and not.

What is data migration as a part of ERP implementation?

Data migration projects can have many forms. The classic form of the data migration is where the new system is to be implemented and needs to be set up with the data from the legacy systems.

The data migration can be defined as a software-supported one-time process migrating data from a source system, which supposed to be shut down, to a target system with a typically different data model. On the other hand, it can be defined as the selection, preparation, extraction, transformation and permanent movement of appropriate and right quality data to the right place at the right time.

There are several approaches for the data migration. A general approach for the data migration can be divided for example into the following elements: extraction including cleansing and preparation, transformation, testing, load, orchestration, fallback, and legacy decommissioning.

This study focused especially on the data migration in SAP ERP implementation projects. Based on the best practices SAP has established the rapid data migration solution which is planned especially on SAP implementations. The solution includes six activities: analyze, extract, clean, validate, load, and reconcile.

It can be concluded that data migration is a crucial part of ERP implementation projects. Issues with the data migration can be costly and can jeopardize the whole implementation project.

What are the barriers for the data quality in the data migration process?

The data migration process has multiple barriers. As the data migration is a part of the ERP implementation project, the barriers for the ERP implementation can be seen as

barriers for the data migration as well. On the other hand, as the data is the key element of the data migration the general barriers for the data quality are applicable also in the data migration process. The barriers in the data migration can be seen as challenges for the data quality.

The data quality barriers in the data migration process can be divided into three groups: data, people, and process related barriers. The data related barriers are directly related to the data itself and the rules, systems, and techniques behind the data. The data related data quality barriers are for example poor quality data in the legacy system, difficulties in data conversion from the legacy system to the target, businesses being unaware of the status and quality of the data in the legacy system, huge volumes of data, and no clear data quality rules or criteria.

The people related barriers are indirectly related to the data but directly related to the people in the data migration process. These barriers include communication, people's roles, human behavior, and people's knowledge and experience related challenges. Good examples of the people related data quality barriers are too little or difficult communication within the project team, unclear roles, ownerships, and responsibilities within the project team, inefficient management of change, no operating master data governance team for the legacy system, and poorly trained or unskillful project team.

The process related data quality barriers are indirectly related to the data but directly related to the processes in the data migration, such as planning and resources. This group combines all the barriers which are not directly related to the data and the people in the data migration. The process related data quality barriers are for example the following: data migration process and the schedule is underestimated, the data migration project is planned poorly, data migration is seen only as a technical process, the data migration project scope is not clear, and documentation and file handling is not disciplined.

This study concentrated especially on the SD master data. However, all the barriers mentioned can also be seen as barriers for the general data quality in the data migration. In addition, all the barriers in the data migration process have some effect on the data quality as the data is the key element in the data migration.

How to improve the master data quality in the data migration of SAP ERP system implementation?

The methods to improve the master data quality in the data migration process were derived with the help of the data quality barriers in the data migration process which were also discussed in the study. The interviews resulted in a long list of methods which

should be taken into consideration in order to improve the master data quality in the data migration. These methods can be divided into three groups: data, people, and process related methods.

The data related methods are directly related to the data itself and the rules, systems, and techniques behind the data. When considering the methods which stood out in the interviews the following distinguished from the list: create unambiguous data collection templates and carry out walkthrough for them, determine unambiguous steps and rules for the data cleansing process, and arrange scheduled checkpoints specifically for the data. In all, the data related methods to improve the master data quality are highly related to the data migration process phases.

The people related methods are indirectly related to the data quality but directly related to the people in the data migration process. As the data related barriers these methods include communication, people's roles, human behavior, and people's knowledge and experience related procedures. The distinguished people related methods to improve the master data quality in the data migration were defining and communicating the data related roles and responsibilities unambiguously, gathering skillful project group, and including and engaging all parties (incl. data people) in the project from the beginning. The people related methods to improve the master data quality in the data migration were more general in nature and can be applicable to any project. However, they have an indirect effect on the data quality.

The process related methods are indirectly related to the data but directly related to the processes in the data migration, such as planning and resources. This group combines all the methods which are not directly related to the data and the people in the data migration. The methods which stood out from the list were the following: plan the data migration project, schedule, roles & responsibilities, etc. carefully and in detailed level and determine rules for disciplined documentation. As well as the people related methods, the process related methods are more general in nature. However, they are still important from the data quality perspective.

In addition to the interviews an additional questionnaire were used to distinguish the most intrinsic methods to improve the master data quality in the data migration process. According to the questionnaire results these most intrinsic methods are:

- Take care of the good engagement with the client
- Define and communicate the data related roles and responsibilities unambiguously (e.g. who extracts, transforms, cleanses, uploads, and sign offs each of the data objects)
- At starting point analyze the status of the data in the legacy systems
- Arrange and execute the data cleansing carefully already in the legacy system

- Create unambiguous data collection templates and carry out walkthrough for them
- Determine SAP rules for data to correspond to the business rules (e.g. mandatory fields)

The first two methods in the list are people related methods and the last four methods are data related methods. The process related methods did not stand out in the questionnaire.

In conclusion, even though the study concentrated on SD master data some of the methods, especially people and process related, are very high-level advice which in any case should be taken into consideration in the data migration process in order to improve the master data quality. The methods don't either give detailed instructions how they can be utilized. Nevertheless, the methods clearly indicate how the master data quality can be improved in the data migration.

7.2 Assessment of the study and further research

In the final analysis, the research answers to the research questions and reaches the goal which was set for the thesis. In addition, the results can be considered reliable as the interviewees were experts in the studied phenomenon. Therefore, the research can be considered as successful.

Considering the chosen research philosophy, approach, strategy, and data collection and analysis methods it can be concluded that they were suitable for the thesis. The hermeneutic philosophy supported the qualitative nature of the research. The inductive and qualitative approach helped to stress the empirical research and make new findings from the interviews. The case study as a strategy also supported the qualitative nature of the research and helped to understand the complex phenomenon in its natural context. Finally, the theme-centered interviews aided to have comprehensive results and the additional questionnaire helped to point out the most intrinsic results.

The theory provided sufficient knowledge for the empirical section of the study. Firstly, the theory worked as a very basis for the empirical research providing the initial information for the data collection. Secondly, it formed a foundation for the research providing the necessary information for the studied phenomenon as a whole. The chapter two introduced the data quality and especially master data quality and what are the effects of poor master data quality for organizations. On the other hand, the chapter three discussed about ERP implementations and data migration processes and what are the barriers for the data migration.

At the end of the day, it turned out to be challenging to find the scientific literature for the studied phenomenon in its natural context. There is lots of literature about master data management and on the other hand ERP implementations and data migrations. However, as this thesis aimed to combine the master data management perspective in the data migration, the suitable literature was hard to find.

The empirical research was conducted as theme-centered one-by-one interviews. In total, nine SAP consultants were interviewed. However, it should be noted that the actual master data management experts or the client side representatives were not interviewed, but the interviewees had a strong experience from SAP implementations and data migration processes. Therefore, they provided valuable and reliable information how the master data should be handled in the data migration process. In addition, they were able to consider the role of the data both managerial, functional, and technical perspectives.

Even though the focus on the SD master data guided the empirical research, the main results were fairly high-level instructions to improve the master data quality in the data migration. However, they were all applicable for the master data and especially SD master data in the studied phenomenon. Therefore, the results can be considered trustworthy. All in all, the results are also applicable for the organizations that are planning to implement SAP ERP system and are concerned about the master data quality.

On the other hand, the results could have been better and more detailed if a second round of interviews had been arranged. Then the results got from the first interviews could have been discussed more closely, which would have resulted more detailed methods to improve the master data quality in the data migration. Then also the focus on the SD master data could have been stressed more. Nevertheless, with the tight schedule it was impossible to execute such a time-consuming iterative process.

This study can form the basis for the future research. Firstly, the results of the study could be used to create a process framework for the data migration where the master data is taken into consideration carefully. The framework could then be utilized in planning and executing the data migration processes in different organizations. In addition, the framework could also guide how the master data should be managed in the data migration. Secondly, the results could be used to develop more detailed step-by-step instructions for the methods. They could advice how the methods to improve the master data quality in the data migration can be put into operation in practice.

BIBLIOGRAPHY

Aaltola, J. & Valli, R. 2007. Ikkunoita tutkimusmetodeihin. 2nd ed. Jyväskyla, PS-kustannus. 243 p.

Alasuutari, P. 2001. Laadullinen tutkimus 2.0. 3rd ed. Jyväskylä, Vastapaino.

Ballou, D. & Pazer, H. 1985. Modeling data and process quality in multi-input multi-output information systems. Management Science, Vol. 31(2), pp. 150-62.

Bobrowski, M., Marré, M., & Yankelevich, D. 1999. Measuring data quality. Universidad de Buenos Aires. Report, 99-002. 11 p.

Butler, T. 1998. Towards a hermeneutic method for interpretive research in information systems. Journal of Information Technology, Vol. 13, pp. 285-300.

Chudy, M. & Castedo, L. 2011. Sales and distribution in SAP ERP: Practical guide. Bonn, Galileo Press. 406 p.

Coffey, A. & Atkinson, P. 1996. Making sense of qualitative data: Complementary research strategies. Thousand Oaks, Sage Publications. 206 p.

Davenport, T.H. 1998. Putting the Enterprise into the Enterprise System. Harvard Business Review, Vol. 76(4), pp. 121–131.

Davenport, T.H., & Prusak, L. 1998. Working Knowledge: How Organizations Manage What They Know. Cambridge, Massachusetts, Harvard Business School Press. 224 p.

Davenport, T.H., Harris, J.G., & Cantrell, S. 2002. The return of enterprise solutions: The director's cut. Accenture Institute for Strategic Change Research Report, 14, 53 p.

Eckerson, W.W. 2002. Data quality and the bottom line. TDWI Report, The Data Warehouse Institute. 32 p.

Feldman, S. & Sherman, C. 2001. The high cost of not finding information. Framingham, Massachusetts, IDS. 10 p.

Haller, K. 2009. Towards the industrialization of data migration: concepts and patterns for standard software implementation projects. In Advanced Information Systems Engineering, pp. 63-78.

Haller, K., Matthes, F. & Schulz, C. 2012. A detailed process model for large scale data migration projects. In 15th International Conference on Business Information Systems, pp. 165-176.

Hancock, D. & Algozzine, R. 2006. Doing case study research: A practical guide for beginning researchers. New York, Teachers College Press. 106 p.

Haug, A., Arlbjørn, J.S., & Pedersen, A. 2009. A classification model of ERP system data quality. Industrial Management & Data Systems, Vol. 109(8), pp. 1053-1068.

Haug, A. & Arlbjørn, J.S. 2011. Barriers to master data quality. Journal of Enterprise Information Management, Vol. 24(3), pp. 288-303.

Haug, A., Zachariassen, F., & Van Liempd, D. 2011. The costs of poor data quality. Journal of Industrial Engineering and Management, Vol. 4(2), pp. 168-193.

Hawking, P., Stein, A. & Foster, S. 2004. Revisiting ERP systems: benefit realization. Proceedings of the 37th Annual Hawaii International Conference on System Sciences, pp. 1-8.

Heracleous, L. 2004. Interpretivist Approaches to Organizational Discourse. In Grant, D., Hardy, C. O. & Putnam, L. (eds.). The Sage handbook of organizational discourse. London, Sage Publications Ltd. pp. 175–192.

Hirsjärvi, S., Remes, P. & Sajavaara, P. 2009. Tutki ja kirjoita. 15th ed. Helsinki, Tammi. 464 p.

Hong, K.K. & Kim, Y.G. 2002. The critical success factors for ERP implementation: an organizational fit perspective. Information & Management, Vol. 40(1), pp. 25-40.

Kanaracus, C. 2014. Cost of troubled SAP project will skyrocket to nearly $1 billion, audit says [e-zine]. Computerworld. Published on 03.10.2014 [accessed on 06.10.2014]. Available: http://www.computerworld.com/article/2691661/cost-of-troubled-sap-project-will-skyrocket-to-nearly-1-billion-audit-says.html

Kanaracus, C. 2013. California ends contract with SAP over troubled IT project [e-zine]. Computerworld. Published on 08.02.2013 [accessed on 06.10.2014]. Available: http://www.computerworld.com/article/2494827/it-management/california-ends-contract-with-sap-over-troubled-it-project.html

Kanaracus, C. 2011. SAP-IBM Payroll System Woes Fouled up Nurses' Pay [e-zine]. PCWorld. Published on 14.07.2011 [accessed on 06.10.2014]. Available: http://www.pcworld.com/article/235738/article.html

Klaus, H., Rosemann, M., & Gable, G.G. 2000. What is ERP?. Information systems frontiers, Vol. 2(2), pp. 141-162.

Knolmayer, G. & Röthlin, M. 2006. Quality of material master data and its effect on the usefulness of distributed ERP systems. Lecture Notes in Computer Science, Vol. 4231, pp. 362-71.

Kogent Learning Solutions, Inc. 2011. SAP SD handbook. Sudbury, Massachusetts, Jones and Bartlett. 632 p.

Krumbholz, M. & Maiden, N. 2001. The implementation of enterprise resource planning packages in different organisational and national cultures. Information systems, Vol. 26(3), pp. 185-204.

Kurbel, K. 2013. Enterprise resource planning and supply chain management: Functions, business processes and software for manufacturing companies. Heidelberg, Springer. 359 p.

Kvale, S. 1996. Interviews: An introduction to qualitative research interviewing. Thousand Oaks, Sage Publications. 326 p.

Lau, L. 2005. Managing business with SAP: planning, implementation and evaluation. Hershey, Pennsylvania, Idea Group Pub. 348 p.

WorldOne Research. 2008. Workplace Productivity Survey [pdf]. LexisNexis. [accessed on 22.12.2014]. Available: http://www.lexisnexis.com/literature/pdfs/LexisNexis_Workplace_Productivity_Survey_2_20_08.pdf

Loshin, D. 2009. Master Data Management. Boston, Morgan Kaufmann, 259 p.

Looi, K. 2009. MDM for customer data optimizing customer centric management of your business. Lewisville, MC Press Online. 135 p.

Madnick, S., Wang, R. & Xian, X. 2003. The design and implementation of a corporate householding knowledge processor to improve data quality. Journal of Management Information Systems, Vol. 20(3), pp. 41-70.

Mandal, P. & Gunasekaran, A. 2003. Issues in implementing ERP: a case study. European Journal of Operational Research, Vol. 146(2), pp. 274-283.

Marsh, R. 2005. Drowning in dirty data? It is time to sink or swim: a four-stage methodology for total data quality management. The Journal of Database Marketing & Customer Strategy Management, Vol. 12(2), pp. 105-112.

Matthes, F., Schulz, C. & Haller, K. 2011. Testing & quality assurance in data migration projects. Proceedings of the 27th IEEE International Conference on Software Maintenance (ICSM), pp. 438-447.

McGilvray, D. 2008. Executing data quality projects ten steps to quality data and trusted information. Amsterdam, Morgan Kaufmann/Elsevier. 352 p.

Monk, E. & Wagner, B. 2009. Concepts in enterprise resource planning. 3rd ed. Australia, Course Technology Cengage Learning. 254 p.

Morris, J. 2012. Practical Data Migration. 2nd ed. Swindon, BCS Learning & Development. 248 p.

Motwani, J., Subramanian, R., & Gopalakrishna, P. 2005. Critical factors for successful ERP implementation: exploratory findings from four case studies. Computers in Industry, Vol. 56(6), pp. 529-544.

Newell, S., Robertson, M., Scarbrough, H., & Swan, J. 2002. Managing Knowledge Work. Basingstoke, Palgrave-Macmillan. 223 p.

Olkkonen, T. 1994. Johdatus teollisuustalouden tutkimustyöhön. 2nd ed. Otaniemi, TKK OFFSET. 143 p.

Olson, J 2003. Data Quality: The Accuracy Dimension. USA, Morgan Kaufmann Publishers. 300 p.

Orr, K. 1998. Data quality and systems theory. Communications of the ACM, Vol. 41, pp. 66–71.

Oyegoke, A. 2011. The constructive research approach in project management research. International Journal of Managing Projects in Business, Vol. 4(4), pp. 573-595.

Padhi, S. 2013. SAP ERP financials quick reference guide SAP ECC 6.0. Dulles, Virginia, Mercury Learning and Information. 525 p.

Phillips, S. 2012. Control your ERP destiny. United States. 253 p.

Pipino, L.L., Lee, Y.W. & Wang, R.Y. 2002. Data quality assessment. Communications of the ACM, Vol. 45(4), pp. 211-218.

Redman, T.C. 1998. The impact of poor data quality on the typical enterprise. Communications of the ACM, Vol. 41(2), pp. 79-82.

Redman, T.C. 2008. Data driven: Profiting from your most important business asset. Boston, Massachusetts, Harvard Business Press. 257 p.

Redman, T.C. 2013. Data's credibility problem. Harvard Business Review, Vol. 91(12), pp. 84-88.

Roebuck, K. 2012. Data Quality High-impact Strategies - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors. Dayboro, Emereo Pub. 690 p.

Sanjongco, M. & Densborn, F. 2013. Migrate your data fast and easily, Rapid data migration: Go live with clean, valid data [pdf]. SAP AG. [accessed on 25.01.2015]. Restricted availability: https://websmp103.sap-ag.de/~sapidb/012002523100012699012014E/DataMigrationOverview.pdf

SAP. 2013a. SAP ERP Sales and Distribution I - Part 1 [e-learning]. SAP Learning Hub. Revised on 12.01.2013 [accessed on 25.02.2015]. Restricted availability: https://performancemanager.successfactors.eu/login?company=learninghub&loginMethod=PWD

SAP. 2013b. Data Migration with SAP Data Services [e-learning]. SAP Learning Hub. Revised on 12.01.2013 [accessed on 28.02.2015]. Restricted availability: https://performancemanager.successfactors.eu/login?company=learninghub&loginMethod=PWD

SAP. 2014a. Rapid data migration: A Fast and Easy Way to Transform and Migrate Your Data Quickly and Cost-Effectively [e-book]. SAP AG. [accessed on 10.01.2015]. Available: http://global.sap.com/community/ebook/2013_11_23102/enUS/index.html

SAP. 2014b. ASAP 8 Methodology for Implementation [www]. SAP AG. [accessed on 17.01.2015]. Available: https://support.sap.com/support-programs-services/methodologies/implement-sap/asap-implementation.html

Saunders, M. & Tosey, P. 2012. The Layers of Research Design. Rapport, (30), pp. 58-59.

Saunders, M., Lewis, P. & Thornhill, A. 2009. Research methods for business students. 5th ed. Harlow, Financial Times Prentice Hall. 614 p.

Shang, S. & Seddon, P.B. 2000. A comprehensive framework for classifying the benefits of ERP systems. AMCIS 2000 Proceedings, 39, pp. 1005-1014.

Souza, C.A., & Zwicker, R. 2009. ERP Systems' Life Cycle: An Extended Version. Encyclopedia of Information Science and Technology, 2nd ed., pp. 1426-1431.

Tayi, G.K. and Ballou, D.P. 1998. Examining data quality. Communications of the ACM, Vol. 41(2), pp. 54-57.

Tashakkori, A. & Teddlie, C. 2010. Sage handbook of mixed methods in social & behavioral research. 2nd ed. Los Angeles, Sage Publications. 893 p.

Tuomi, J. & Sarajärvi, A. 2002. Laadullinen tutkimus ja sisällönanalyysi. Helsinki, Tammi. 158 p.

Umar, A., Karabatis, G., Ness, L., Horowitz, B., & Elmagardmid, A. 1999. Enterprise data quality: A pragmatic approach. Information Systems Frontiers, Vol 1.(3), pp. 279-301.

Umble, E.J., Haft, R.R. & Umble, M.M. 2002. Enterprise resource planning: Implementation procedures and critical success factors. European Journal of Operational Research, Vol. 146(2), pp. 241-257.

Vaus, D. 2002. Surveys in social research. 5th ed. London, Routledge. 379 p.

Wagner, J. 2013. ERP: 68 most asked questions - what you need to know. Emereo Publishing. 64 p.

Wand, Y. & Wang, R.Y. 1996. Anchoring data quality dimensions in ontological foundations. Communications of the ACM, Vol. 39(11), pp. 86-95.

Wang, R.Y., Storey, V.C. & Firth, C.P. 1995. A framework for analysis of data quality research. IEEE Transactions on Knowledge and Data Engineering, Vol. 7(4), pp. 623-40.

Wang, R.Y. & Strong, D.M. 1996. Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems, Vol. 12(4), pp. 5-34.

Wu, B., Lawless, D., Bisbal, J., Richardson, R., Grimson, J., Wade, V. & O'Sullivan, D. 1997. The butterfly methodology: A gateway-free approach for migrating legacy information systems. Proceedings of the Third IEEE International Conference on Engineering of Complex Computer Systems (ICECCS), pp. 200-205.

Xu, H., Horn Nord, J., Brown, N., & Daryl Nord, G. 2002. Data quality issues in implementing an ERP. Industrial Management & Data Systems, Vol. 102(1), pp. 47-58.

Yin, R. 2009. Case study research: Design and methods. 4th ed. Los Angeles, Sage Publications. 285 p.

Yin, R. 2011. Qualitative research from start to finish. New York, Guilford Press. 369 p.

Zack, M.H. 1999. Managing codified knowledge. Sloan Management Review, Vol. 40(4), pp. 45-58.

APPENDIX 1: INTERVIEW PROTOCOL

Interview – Master of Science Thesis

Preface

- Research question: How to improve the master data quality in the data migration of SAP ERP system implementation?

- The interview is anonymous.
- The interview will be recorded if the interviewee gives his/her assent to it.
- The interviewee does not have to share any confidential information, e.g. names, organizations, etc.
- This interview does not concentrate on specific organizations, customer projects or cases, i.e. the research concentrates on the studied phenomenon in general.
- The most important goal is to discuss about the problems and issues in data migration, reflect the problems to data quality, and finally get recommendations and guidelines how to improve data quality in data migration.
- Focus:
    - Sales & Distribution related master data
    - Legacy system: any system
    - Target system: SAP ERP

Background questions

- Current job/ project role
- In which of the following groups you would place yourself: managing, technical, functional, other
- Relevant experience by years

Theme 1. Data quality

- How would you define data quality?
- What do you think about the thought: “Data quality is overrated.”?
- How would you emphasize the importance of data quality?

Theme 2. Perfect data quality

- Is perfect data quality required?

- If perfect data quality is not required, what is?
- Are some data quality issues acceptable?

Theme 3. Data quality problems

- What would you say are the most common problems or issues with the data quality?
- Are some problems harder to handle than others?

Theme 4. Data migration process

- How data quality is usually taken into consideration in data migration process?
- How data quality is usually managed in data migration?
- What is the role of business?
- What is the role of the migration team?

Theme 5. Problems in the data migration process

- What kind of data migration process is unsuccessful?
- What usually goes wrong?
- What kind of problems you've confronted in data migration?
- Who takes responsibility of the data quality problems in data migration?
- Are there differences between different cultures?

Theme 6. Data quality barriers in the data migration (SD Master data)

- Thinking about the problems in data migration which ones are also barriers for the data quality?
- What are the main reasons for the barriers?
- Are some data objects more problematic than others?
- Do you think that the systems support the high quality data enough?

Theme 7. Improving the data migration process

- How would you describe a successful data migration process?
- When considering the problems and barriers how they could be encountered/ solved?
- How to improve data migration process?

Theme 8. Improving the data quality in the data migration (SD Master data)

- By which methods the data quality could be improved in the data migration?
- How the data quality management should be organized in the data migration?
- Which steps in the data migration are the most important from the data quality perspective?

- Who should be responsible for the data quality?
- How to accelerate the data cleansing and purging process?
- How to measure the data quality?

Theme 9. The lessons learned/ Golden rules

- Do you have any lessons learned or golden rules for the future data migration processes?

Theme 10. Prospects/ Future

- What do you think about the prospects for the ERP implementation and data migration?
- How it will effect on the data quality?
- Do you think that the data quality problems still remain same?

APPENDIX 2: ADDITIONAL QUESTIONNAIRE

Improving master data quality in data migration of ERP implementation project

The questionnaire includes only one compulsory question where I'd like you to select the most intrinsic methods to improve the master data quality in the data migration process of an ERP implementation. The methods are collected from the interviews and listed in randomized order below.

Please select the most intrinsic methods to improve the master data quality in the data migration process (select 5 choices).

- ☐ Comply with SAP best practices always when applicable
- ☐ Improve and clarify the quality of the metadata
- ☐ Determine SAP rules for data to correspond to the business rules (e.g. mandatory fields)
- ☐ Determine unambiguous steps and rules for the data cleansing process (e.g. what data is obsolete, how to check duplicates, etc.)
- ☐ Take care of change management
- ☐ Emphasize the partnership with the client
- ☐ Emphasize and facilitate the communication within the project team
- ☐ Separate and emphasize the data quality improvement process as a part of the data migration process
- ☐ Determine rules for disciplined documentation (including file handling, naming conventions, change log, etc.)
- ☐ Do multiple test uploads for the data to different systems
- ☐ Arrange the introduction and boarding to the project team carefully
- ☐ Require a sign off for the data after every test upload
- ☐ Gather skillful project group (functional, technical and business knowledge, etc.)
- ☐ Release the project people from their day job to the project
- ☐ Define and communicate the data related roles and responsibilities unambiguosly (e.g. who extracts, transforms, cleanses, uploads, and sign offs each of the data objects)
- ☐ Create unambiguous data collection templates and carry out walkthrough for them
- ☐ Raise the importance of good quality data and get buy in from the business at the early stages of the project
- ☐ Arrange and execute the data cleansing carefully already in the legacy system
- ☐ Arrange a solution walkthrough for the customer of tools that can be appropriate for the project
- ☐ Reduce the amount of the migrated data considerably by different criteria (do not migrate redundant data)
- ☐ Define the parameters for the data quality (how the data quality can be measured and proven in the project)
- ☐ Arrange the cutover simulation where data is also tested

- In the user acceptance testing (UAT) test also data quality
- Determine clear reconciliation and sign off methods for the data (use tools if available)
- Minimize the customization and workarounds in SAP (respect standard SAP solutions)
- Upload the whole data scope also to the test system and test complete data sets
- Manage the general (a-level) data centrally in one place (creation, maintenance, etc.)
- Arrange scheduled checkpoints specifically for the data
- Define unambiguous rules for the required content and format of the data (e.g. capital letters, phone number format, etc.)
- Clarify the goal in the beginning for the project members
- Arrange a separate system environment (playground) for data related tests
- Always clarify the business reasons behind the data fields and values for the customer
- Communicate the project scope, phases, schedule, roles etc. for all project members
- In test uploads and sign offs concentrate on validating only the most important data (e.g. most important customers)
- Plan the data migration project, schedule, roles & responsibilities, etc. carefully and in detailed level
- Utilize the tools and applications available
- At starting point analyze the status of the data in the legacy systems
- Include and engage all parties (incl. data people) in the project from the beginning
- Take care of the good engagement with the client